



# DESIGN AND IMPLEMENTATION OF HIGH SPEED MULTIMEDIA NETWORK

By

YEUNG CHUNG TOA

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF PHILOSOPHY

DIVISION OF INFORMATION ENGINEERING

THE CHINESE UNIVERSITY OF HONG KONG

JUNE 1994

UL

Meis  
TK  
5105.5  
Y48  
1994

# Acknowledgement

It is my pleasure to acknowledge my debt to many people who have contributed to this research project. Their suggestions have given me many good ideas in correcting the mistakes and improving the prototype.

I would like to express my warmest gratitude to Dr. Cheung Kwok-Wai, who is my supervisor for this project. His sincere help and advice have made not only me, but the whole CUM LAUDE NET work team successful. During the course of study and frequent discussions, I have gained a lot of knowledge about high speed multimedia network.

I would also like to thank Mr. Alex Siu and Mr. Y.W. Lee for their technical advices, and the CUM LAUDE NET team member, namely, W.K.Lam, C.H.Chan, C.W.Chung, C.Lau, K.K.Lau, O.Soo, C.K.Tong, K.F.Wong, and K.F.Yuen, for their contribution in giving me valuable ideas in the design of the router.

# Abstract

CUM LAUDE NET is a high-speed (Gb/s) multimedia integrated network being designed and prototyped at the Chinese University of Hong Kong. Current facilities include a fully operational video and voice conferencing utility, a gateway to the public switched telephone network, voice mail services as well as other internet services. This thesis will describe the architectural design and hardware implementation of CUM LAUDE NET. In particular, the design and implementation of a low-cost, 100-Mb/s network router based on fast packet routing and a 100-Mb/s network interface card for local host will be described.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Bandwidth required by multimedia applications . . . . .	1
1.2	Real-time requirement . . . . .	2
1.3	Multicasting . . . . .	2
1.4	Other networks . . . . .	3
1.5	Overview of CUM LAUDE NET . . . . .	5
1.5.1	Protocols . . . . .	7
1.5.2	Network Services . . . . .	8
1.6	Scope of the Thesis . . . . .	9
<b>2</b>	<b>Network Architecture</b>	<b>11</b>
2.1	CUM LAUDE NET Architectural Overview . . . . .	11
2.2	Level One Network Architecture . . . . .	12
2.3	Level-One Router . . . . .	14
2.3.1	packet forwarding . . . . .	14
2.3.2	packet insertion . . . . .	15
2.3.3	packet removal . . . . .	15
2.3.4	fault protection . . . . .	15

2.4	Hub . . . . .	16
2.5	Host & Network Interface Card . . . . .	17
<b>3</b>	<b>Protocol</b>	<b>19</b>
3.1	Design Overview . . . . .	19
3.2	Layering . . . . .	20
3.3	Segment, Datagram, and Packet Format . . . . .	21
3.3.1	IP/VCI field . . . . .	23
3.4	Data Link . . . . .	23
3.4.1	byte format and data link synchronization . . . . .	23
3.4.2	access control byte . . . . .	24
3.4.3	packet/frame boundary . . . . .	26
3.5	Fast Packet Routing Protocol . . . . .	26
3.5.1	Level-2/Level-1 Bridge/Router . . . . .	27
3.5.2	Level-1 Hub . . . . .	29
3.5.3	Local Host NIC . . . . .	29
3.6	Media Access Control Protocol I : ACTA . . . . .	30
3.7	Media Access Control Protocol II: Hub Polling . . . . .	34
3.8	Protocol Implementation on CUM LAUDE NET . . . . .	36
<b>4</b>	<b>Hardware Implementation &amp; Performance of Routers and NIC</b>	<b>40</b>
4.1	Functionality of Router . . . . .	40
4.2	Important Components Used in the Router Design . . . . .	43
4.2.1	TAXI Transmitter and Receiver . . . . .	43
4.2.2	First-In-First-Out Memory (FIFO) . . . . .	44
4.3	Design of Router . . . . .	45

4.3.1	Version 1	45
4.3.2	Version 2	47
4.3.3	Version 3	50
4.4	Lessons Learned from the High Speed Router Design	57
<b>5</b>	<b>Conclusion</b>	<b>61</b>
	<b>Bibliography</b>	<b>63</b>

# Chapter 1

## Introduction

In this era of information, the demand for communication bandwidth continues to explode in the society. In the past twenty years, people were quite satisfied about the performance of 10-Mb/s Ethernet network [1] connected to public data networks like X.25 [2]. This was because most applications like E-mail and file transfers were non-real time processes with a relatively small volume which could be easily served by these networks. With increasing number of users and with more and more bandwidth-consuming applications such as multimedia running on the network, the available network bandwidth is under huge demand.

### 1.1 Bandwidth required by multimedia applications

Multimedia [3, 4] is one of the most popular topics at present. Multimedia is the integration of continuous media such as voice and video with traditional data like image and text. The bandwidth needed for voice is relatively small.



For uncompressed PCM voice, the bandwidth required is 64 kb/s. However, the bandwidth needed for continuous video is very large. For an uncompressed full motion video at 30 frames per second using a resolution of 320 x 200 pixels and 256 colors, the data rate required is 1.92 MB/s (15.36 Mb/s). It can be seen that this is already beyond the ability for Ethernet to carry.

## **1.2 Real-time requirement**

Besides the considerations on bandwidth, the ability for a network to provide real-time data transfer is another important issue [5]. Take voice transmission on Ethernet as an example, when the traffic loading on Ethernet is very low, real-time voice transmission usually presents no problem even to the CSMA/CD protocol. However, when the loading on Ethernet increases, the chance of packet collision is so high that voice packet would not be able to be delivered to its destination in real-time even if the bandwidth of Ethernet is sufficient to carry the voice data. The delay jitter characteristics of a network is very important for real-time multimedia applications.

## **1.3 Multicasting**

Furthermore, there are many multimedia applications like video conferencing that require multiple drop destinations. That is to say, a single data packet transmitted by a station could be received by more than one destinations. The Ethernet protocol is not suitable for such multicasting applications. If we attempt to use Ethernet's broadcasting mode to perform packet multicasting, the

overall network performance will be degraded severely since all stations connected to the network have to process all the broadcast packets whether the packet is addressed to the station or not.

## **1.4 Other networks**

In recent years, many higher speed networks have been proposed for carrying multimedia traffic, such as ATM [6, 7, 8], FDDI [9, 10], the Cambridge Ring [11] and the HP Hangman [12].

The Asynchronous Transfer Mode (ATM) is a very promising approach to multimedia networking. It has been standardized by international standard organizations [7, 8]. ATM employs fixed-size cells (53 bytes) with a 5-byte header. Because of the small cell size, virtual circuit addressing and simple protocol processing must be used. This implies a very heavy burden and demand for a very sophisticated switching system design. At the moment of this writeup, large ATM switching systems are still not yet available, impeding the large scale deployment of ATM technology by service operators such as cable companies and telephone companies. Furthermore, the network management, signalling and control of ATM networks are not yet standardized. As it is, ATM is still the network of the future.

Fiber Distributed Data Interface (FDDI) is a 100-Mb/s token ring that uses multimode optical fiber as transmission medium. It has become a high speed local area network standard. FDDI is not suitable for maintaining a continuous, constant-data-rate connection between two stations because the synchronous data transfer of FDDI-I does not guarantee a uniform data stream. Such a



constant data stream is typical of circuit-switched application, such as digitized voice or video. FDDI-II is an upward-compatible extension to FDDI-I that supports circuit-switched traffic in addition to the packet mode traffic supported by the FDDI-I. Although FDDI-II is a possible solution to multimedia applications, the price is so too high that it is not yet practical to use FDDI-II to connect every station other than the backbone network itself. Due to the cost effectiveness, Ethernet is still the most popular network used in the market.

The HANGMAN Gb/s network [12] is a dual-ring network. A 1-Gb/s three-node network has been constructed at the Bristol Laboratories. A video conferencing demonstration consists of two standard video cameras attached to personal computers. Video signal are digitized by custom boards within PC and format it into packets. It is then passed to the interface board and transmitted across the network. The destination PC receives the packets, recovers the video information, and displays it on the screen. The HANGMAN network is a very successful multimedia network. However, each node on the HANGMAN network can only support a single host, making each host rather costly.

In order to overcome this limitation, a high speed and affordable multimedia network is designed and prototyped. It is called the CUM LAUDE NET [13]. The CUM LAUDE NET is a large group effort at the Chinese University of Hong Kong to prototype a multi-gigabit/sec multimedia integrated network. It started in early 1993 and is scheduled to be completed in summer 1995. The phase I objective is to demonstrate a practical, low-cost and high-speed integrated network that can provide real-time video and voice conferencing. The physical layer design is aimed to be as general as possible, so the technology can be made compatible to ATM, FDDI or any other standards easily.



This thesis will focus on the design and prototyping of the Phase I 100-Mb/s network. The prototype in Phase I consists of a three-node backbone connected in a fault tolerant dual-ring. Each node is able to route 100 Mb/s at each input and output. Each node can allow up to 16 local host attachments to reduce the network cost. The architecture is scalable and can be connected to a multi-gigabit/sec dual-ring backbone in the phase II target. At the time of this writeup, a 100-Mb/s, 3-node multimedia integrated network has been designed and constructed. It is connected by multimode optical fibers and provides good quality real-time video and voice conferencing very inexpensively.

## **1.5 Overview of CUM LAUDE NET**

CUM LAUDE NET is a multi-gigabit/sec hierarchical dual ring network. The prototype that is being constructed consists of two hierarchies as shown in figure 1.1. The level-2 hierarchy is a 1-Gb/s backbone connected in a fault tolerant dual-ring. It aims at providing high-speed, real-time multimedia services for a metropolitan area. Each node will be able to route one gigabit/sec bandwidth at each input and output. Assuming uniform traffic, the final backbone dual-ring network will have an aggregate capacity close to 8 gigabit/sec.

Each level-1 hierarchy is a 100-Mb/s dual ring network, aims at providing the same services for a more local area environment. Different hierarchies are connected by bridges / routers, whose function is to pass packets from one hierarchy to another. At the time of this writeup, an Ethernet gateway which connects CUM LAUDE NET and our departmental network has been successfully constructed. Stations on CUM LAUDE NET can be connected to any



workstations in our department by using TELNET and FTP. This thesis will concentrate on the design and prototype of the level-1 dual-ring network.

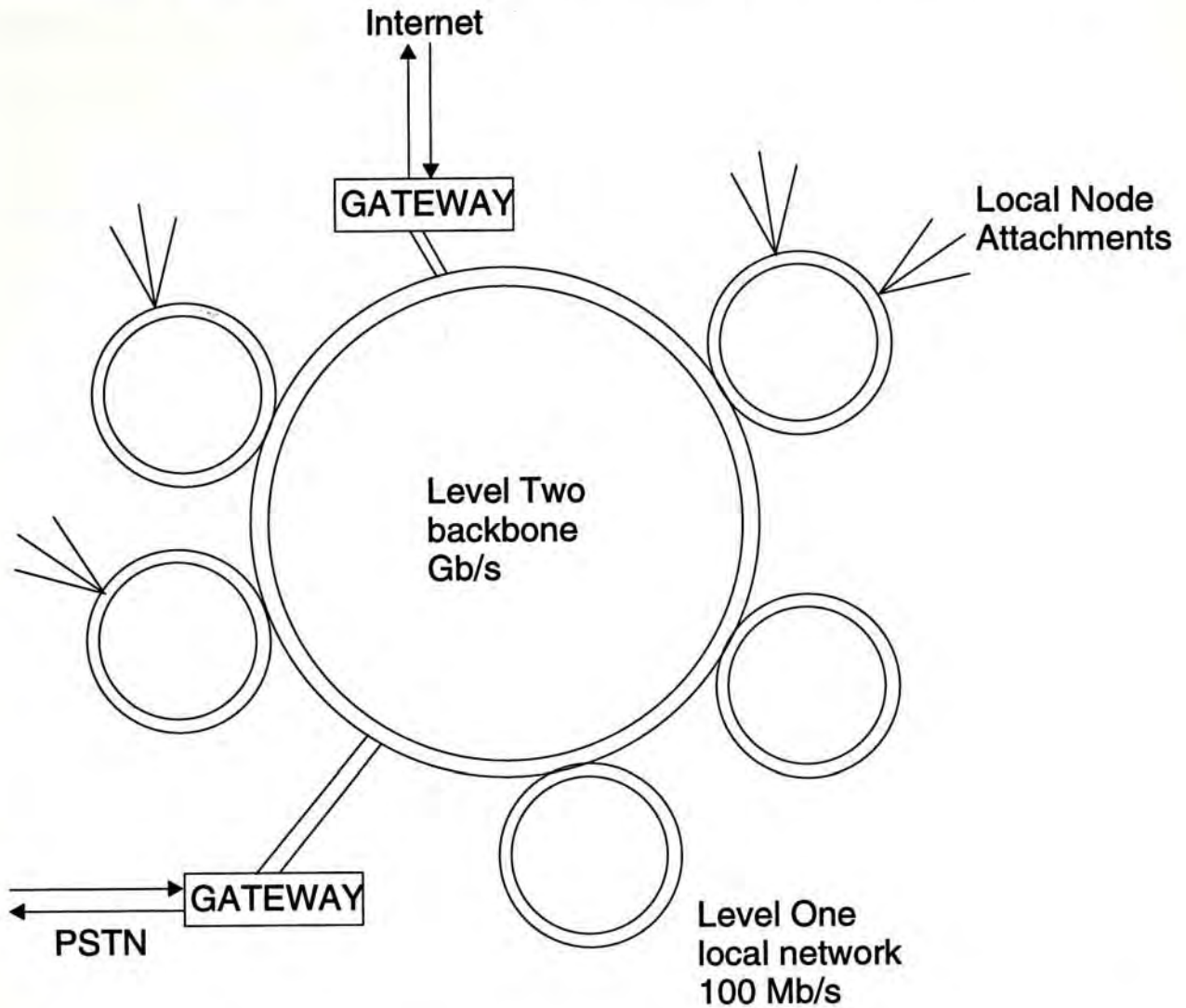


Figure 1.1: CUM LAUDE NET Network Architecture

Each node in the level-1 hierarchy is either a local host (user terminal) or a hub (which serves as a concentrator / server to a number of local hosts). Each local host is equipped with a network interface card (NIC) whose function is to process packets addressed between the network and the host. Due to the uniform design of different hierarchies, the hardware and software design of the bridges / routers are very much simplified.

The CUM LAUDE NET is designed around a fault-tolerant dual-ring topology because the ring topology has several unique advantages over a centralized switching hub topology [14, 15]. The linear topology allows reserved service guarantee and fair sharing of bandwidth among all nodes. The distributive, sequential arrangement of the nodes also facilitates real-time protocol implementation. Distributive packet routing simplifies packet processing and introduces little packet delay. Another advantage of the linear topology is that it reduces the problems due to network congestion and complexities in control and management.

Even though many bus/ring network protocols has been proposed in the past [1, 11, 16, 17, 18, 19, 20], they are not suitable for high-speed multimedia integrated networking either because of the limited throughput, inability to guarantee real-time services, confinement to local area service, or heavy overhead for supporting multimedia services.

### **1.5.1 Protocols**

CUM LAUDE NET is designed to support high-speed, real-time multimedia services with maximum compatibility to IP-based networks. The motivation is that Internet is a worldwide network service, has a very broad user base, and yet, Internet does not support real-time multimedia services. Thus, our design could provide an easy upgrade for IP-based networks to the future multimedia networks.

In order to achieve these goals, we have decided to use fast packet routing and integrated networking technology that employs:

1. fixed size IP datagrams / Fast Packet Routing (FPR) packets (576 / 582



octets)

2. FPR in the MAC and network layer
3. Direct IP addressing in the transport and routing of IP datagrams
4. Connectionless delivery of packets

In the Fast Packet Routing (FPR) layer, a layer which combines the MAC layer with some of the network layer functions, the IP datagrams are encapsulated by a fixed-size header and trailer. All the routing information is also available in the header. This allows each router to perform fast packet routing efficiently and simplifies the gateway design between CUM LAUDE NET and Internet. This feature is considered to be better than AAL3/4 or AAL5 [21] in ATM network.

A novel network protocol ACTA (Adaptive Cycle Tunable Access) [14, 15] is implemented in each Level-2 and Level-1 dual-ring hierarchy. It is a slotted access protocol. Fair access is achieved by limiting the number of empty slots occupied by each cycle. The cycle length is adjusted to reduce the packet latency and to increase the overall throughput.

### **1.5.2 Network Services**

The protocol used in the CUM LAUDE NET can support two kinds of services, namely, the unreserved service and the reserved service. The unreserved service provides a mean of fair network access by limiting the number of slots access during a cycle. The reserved service provides a means of slots reservation to guarantee the bandwidth of real-time data transmission. Both services can provide an ideal and fair access platform for multimedia applications.

CUM LAUDE NET has already been connected to Internet, and all standard Internet services like electronic mail (SMTP), remote login (TELNET), and file transfer (FTP) have been supported. The user interface is the industry-standard X-windows. The network has also been connected to the public telephone network through a T-1 gateway, thus allowing CUM LAUDE NET users to call up any telephone users and to send / receive voice mails or other customized services over a computer network.

## **1.6 Scope of the Thesis**

This thesis will focus on the design and implementation of CUM LAUDE NET level-1 hierarchical dual-ring network running at 100-Mb/s on each ring. The main contributions of this thesis are:

1. design and implementation of a 100-Mb/s, low-cost network interface card (NIC),
2. design and implementation of a practical, low-cost level-1 dual-ring router,
3. a practical implementation of ACTA protocol on the network router.

Many different versions of network interface card has been designed and tested. The lessons learned from these different versions will also be discussed.

This thesis consists of five chapters. Chapter 1 is a general introduction of the CUM LAUDE NET. Chapter 2 describes the network architecture of CUM LAUDE NET. Chapter 3 describes the protocol used in CUM LAUDE NET. The theory of ACTA protocol and the practical implementation of ACTA protocol implemented on level-1 route will also be described. Chapter 4 describes different



versions of network interface card constructed and the final version of router. A solution to the cost effective, high speed level-1 hierarchical dual-ring network at 100-Mb/s on each ring is suggested. The performance measurement of the network interface cards and routers will also be discussed. Chapter 5 concludes what has been done and summarizes our experience in the hardware design.

# Chapter 2

## Network Architecture

### 2.1 CUM LAUDE NET Architectural Overview

CUM LAUDE NET is a multigigabit per second hierarchical dual ring network. The network architecture is shown in Figure 1.1. Only two hierarchies are considered for the time being. Each router on the level-2 backbone is connected by a 1-Gb/s link in each direction. As can be seen from the figure, many level-1 dual-ring local networks can be attached to the level-2 backbone through the level-2 routers. These routers are capable of transferring data packets from the level-1 local networks to the level-2 backbone network and vice versa. They can also forward any data packets from among the routers.

Different networks like Ethernet and FDDI can be connected to the CUM LAUDE NET through a gateway. For example, a gateway connecting the CUM LAUDE NET and our departmental Ethernet has been successfully completed. Stations on CUM LAUDE NET can connect to any station on our department's network through this gateway. Internet services like TELNET and FTP have

been supported.

## 2.2 Level One Network Architecture

The level-1 network is a 100-Mb/s dual-ring network intended for local area applications. At 100-Mb/s, multimedia applications like video conferencing can easily be supported. Besides, the transmitter and receiver chipset at 100-Mb/s is readily available on the market.

The link between any two neighboring routers can be seen as a point to point transmission link. Each packet is a self-routing IP datagram, i.e., the packet header contains enough information for routing the packet to its destination.

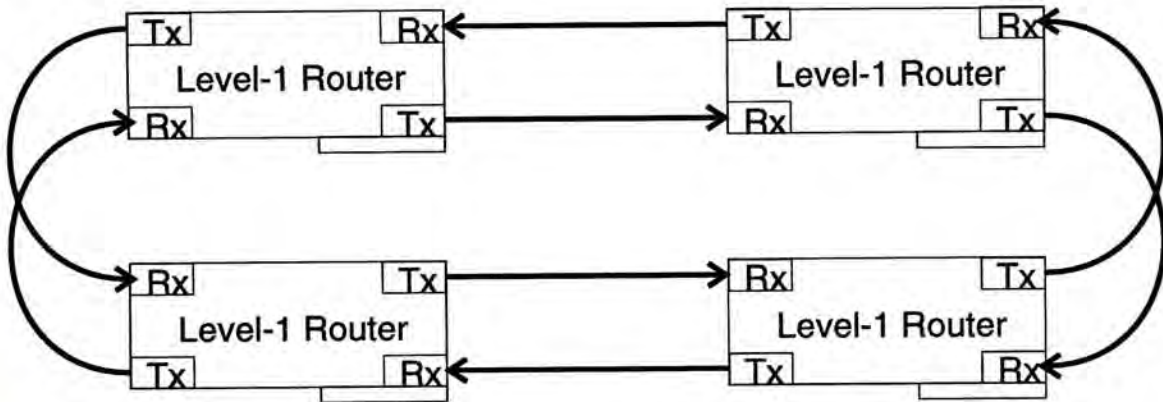


Figure 2.1: Connections of Level One Routers

Figure 2.1 shows a scenario for connecting four level-1 routers. The level-1 routers are connected to one another to form a closed looped-bus in a physical dual-ring. Both rings are unidirectional running opposite to each other.

The physical link is based on byte-oriented transmission. Each byte is encoded by two 4B5B or 5B6B encoders before they are transmitted. Each router acts as an active repeater to transfer data from one router to the next. The



receiver of the next router then decodes and regenerates the serial bit stream and retransmits (or forwards) them to the next router on the same ring.

Among these routers, one of the routers acts as the head of bus and performs the erasure node function. Fixed-sized empty time slots are generated by the head of bus while the returned (occupied or empty) slots are removed by the erasure node.

Since each host station can rarely transfer data at 100-Mb/s, for cost effectiveness consideration, each router can connect up 16 host stations. Each host station is equipped with a network interface card. The network interface cards are connected to the router through a hub. The relationship of the level-1 router, the hub and the level-0 network interface cards is shown in Figure 2.2.

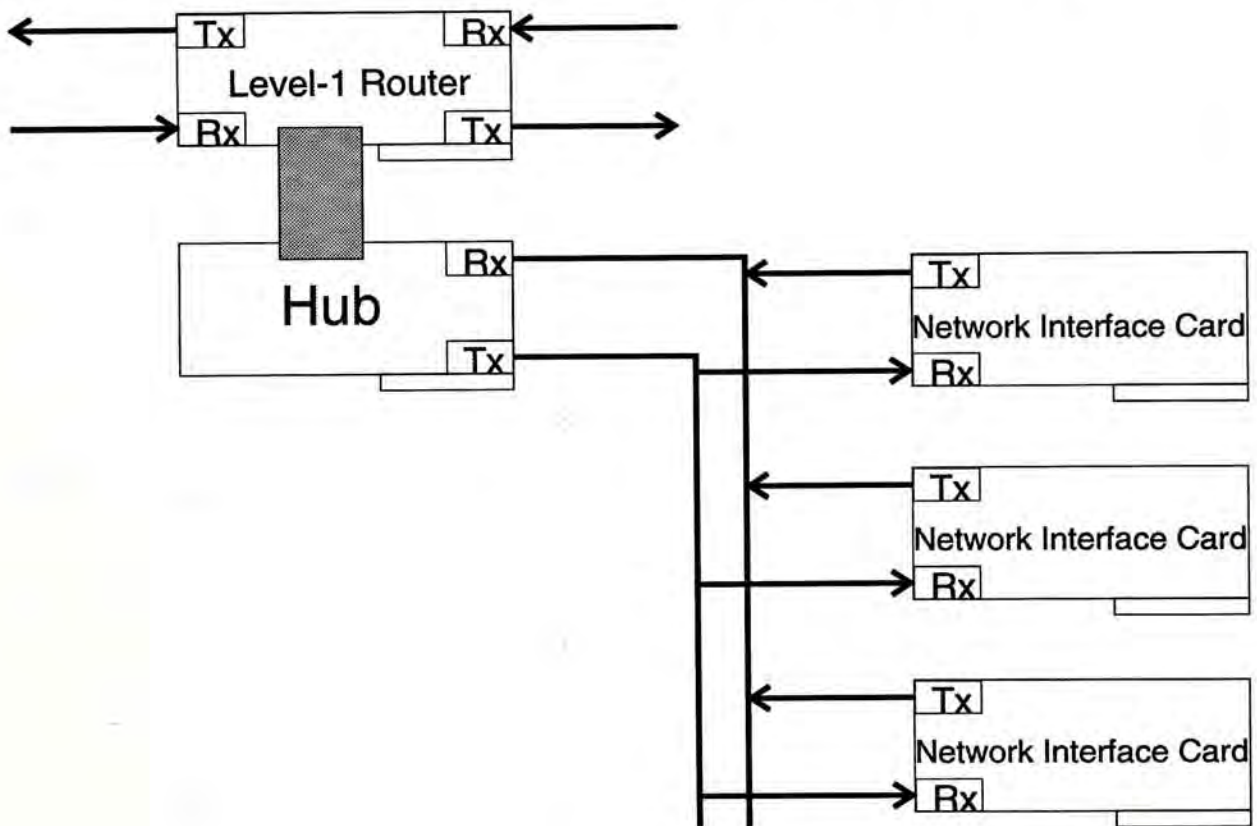


Figure 2.2: Connection of Level-1 Router, Hub and Level-0 Network Interface Card



## **2.3 Level-One Router**

Each level-1 router on the dual ring performs three main functions: packet forwarding, packet insertion (access control), and packet removal. Each router, besides acting as an active repeater, also serves as a network access point for the hub / host stations to gain access to the network.

There are two possibilities for each time slot arrived at the router, namely, an empty time slot or an occupied slot. If an occupied slot arrived, there are three procedures to be performed, namely, packet header extraction, packet identification and routing action. During header extraction, the router will extract the header and address field of an arriving packet. In the packet identification stage, the router attempts to identify the packet type (point to point or multicast) and destination by reading the destination address of the incoming packet in order to perform the appropriate routing action. In the packet routing stage, the action will depend on the protocol used. For the ACTA protocol, an empty slot may be removed from the ring and replaced by a packet-filled slot sent by an attached host. A packet may also be routed to the attached hosts.

### **2.3.1 packet forwarding**

Whether the packet is addressed to the connected host station or not, the packet will be forwarded by the present router to the next one downstream on the ring. If the destination address of the packet matches any of the host stations attached to the router, the packet will be copied from the router to the hub and then to the host station.

### 2.3.2 packet insertion

The packet insertion function of the router allows packets from the hub / host stations to be transmitted to the network. Flow control is necessary to limit the number of packets being sent to the network by the hub. The ACTA protocol is used by the level-1 network to guarantee fair network access among the level-1 routers. The ACTA protocol will be discussed in the next chapter.

The packet insertion function can be carried out only when there is an empty packet arrived at the router and the connected hub / host stations have something to transmit. The empty packet will be removed from the ring and at the same time the packet from the host is substituted.

### 2.3.3 packet removal

Packet removal is only done at the erasure node. The erasure node stops any packet recirculating into the ring and at the same time it performs bandwidth balancing by adjusting the cycle length of the ACTA protocol.

### 2.3.4 fault protection

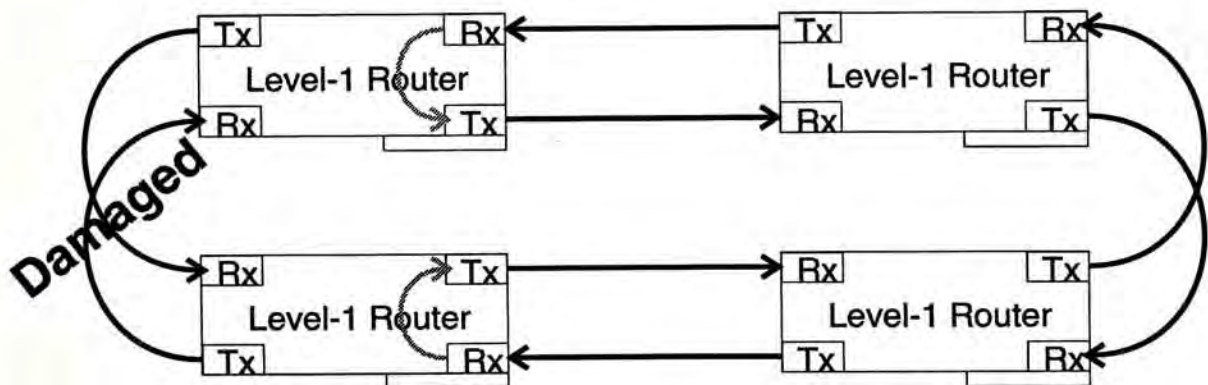


Figure 2.3: Configuration on Link Fault



The dual-ring architecture of CUM LAUDE NET provides a protection to link faults. As shown in Figure 2.3, if a link is damaged, the routers (either level-2 or level-1) will reconfigure their connectivity into a single ring. In this way, packets can still be transmitted to their destination. The penalty is a reduced throughput because there is only one single ring in operation. If there is a router failure, the neighbors of the failed router will be reconfigure so that a single ring is still formed as shown in Figure 2.4. The network is still operational. This is the advantage of having a dual ring architecture to achieve fault tolerant characteristics.

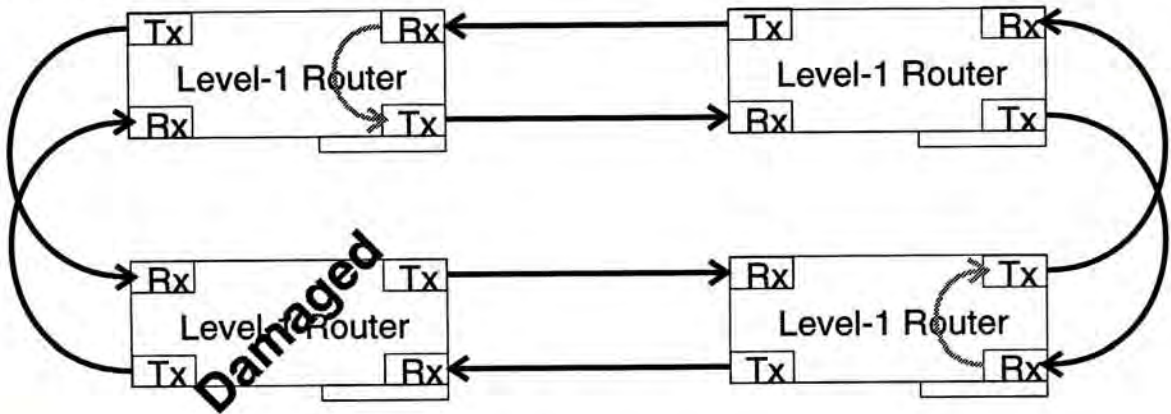


Figure 2.4: Configuration on Router Fault

## 2.4 Hub

Each router is capable of handling 100-Mb/s data on each ring. For the dual-ring architecture, the router is capable of handling at least 200-Mb/s data. As each host is unable to transfer data at the speed of router, to improve the cost effectiveness, multiple hosts can be connected to a router. In this way, the 200-Mb/s bandwidth can be shared among these connected hosts. The cost of the router can also be shared among the number of connected hosts.



The interface between the router and the hosts is called the hub. The hub acts as a concentrator to allow multiple hosts to share the bandwidth of the router. To send packets from the router to the hosts, the packets arrived at the hub are simply broadcasted to the hosts. Only those packets that have destination addresses matching the hosts connected to the hub are broadcasted. To receive packets from the hosts to the router, a round-robin hub-polling scheme is employed to prevent packet collision. Hosts must wait for the token before it can transmit any packet to the hub. The use of the hub polling scheme can ensure that only one host can transmit at a time. The connection among the hub and the hosts are shown in figure 2.2.

## **2.5 Host & Network Interface Card**

In our demonstration network, the hosts are 80486DX-2 66MHz PCs running a UNIX compatible operating system. Each host is equipped with a network interface card which connects the host to the hub. The network interface card needs to perform packet reception and transmission when the host is being polled.

As described in the previous section, the hub will broadcast received packets to all connected hosts. It is the responsibility of each host to extract the appropriate packets with the correct destination address. If a packet's destination address matches the host address (IP or VCI), the packet will be taken by the host for further processing, otherwise, the packet will be ignored by the host.

To transmit packets from the host to the hub, the packet is first written to the network interface card. The network interface card always monitors the

polling sequence of the hub. When the polled address matches its own address, it will respond to the hub and send out the packet. The polling sequence of the hub is transparent to the host.

The network interface card must also perform interrupt to the host after a packet is received. This is done as follows. When the host encounters the trailer of a packet, the network interface card will make an interrupt request to the host to indicate that a complete packet has been read. Thus, the host does not need to know the exact length of the data packet. The host can compare the packet length field of the packet with the actual length read. If there is a length error, the packet will be discarded.

In summary, the current design allows variable packet size to be used, ATM compatibility, supports multicast, and real-time traffic.

# Chapter 3

## Protocol

### 3.1 Design Overview

CUM LAUDE NET is designed to support high-speed, real-time multimedia services with maximum compatibility to IP-based networks. The motivation is that Internet is a worldwide network service, has a very broad user base, and yet, Internet does not support real-time multimedia service. Thus, our design could provide an easy upgrade for IP-based networks to the future multimedia networks.

In order to achieve these goals, we have decided to use fast packet routing and integrated networking technology that employs :

- fixed size IP datagrams/FPR packets (576/582 bytes)
- fast packet routing (FPR) in the MAC and network layer
- direct IP addressing in the transport and routing of IP datagrams
- connectionless delivery of packets



In the Fast Packet Routing (FPR) Layer, which is combining the MAC layer and some of the network layer functions, the IP datagrams are encapsulated by a fixed-size header and trailer and all routing information is available in the header. This allows each router to perform fast packet routing efficiently and simplifies the gateway design between CUM LAUDE NET and Internet. We consider this to be a better feature than AAL3/4 or AAL5 in ATM networks.

A novel network protocol ACTA (Adaptive Cycle Tunable Access) [14] is implemented in each Level-2 and Level-1 dual-ring hierarchy. Fair access is achieved by limiting the number of empty slots occupied by each router on each cycle. The cycle length is adjusted to reduce the packet latency and to increase the throughput.

## **3.2 Layering**

The CUM LAUDE NET protocol layering is shown in Figure 3.1. The protocol is basically an extension of the Internet protocol suite, and is designed to have maximum compatibility with IP. Since TCP is not suitable for real-time multimedia application, a new video and voice transport protocol, VVTP, which is more suitable for carrying real-time video and voice is designed.

VVTP is similar to UDP or TP0, but it has a fixed size, and does not perform acknowledgement, error detection/correction or error retransmission. The design decisions are based on the needs for fast packet routing and the fact that many TCP functions like acknowledgement, error detection/correction and error retransmission are too slow or unnecessary for real-time video and voice applications. The VVTP fragment size is chosen to be 552 bytes. The

corresponding IP datagram/encapsulation have a size of 576 bytes, which is the recommended size that can be handled by Internet networks and gateways without fragmentation.

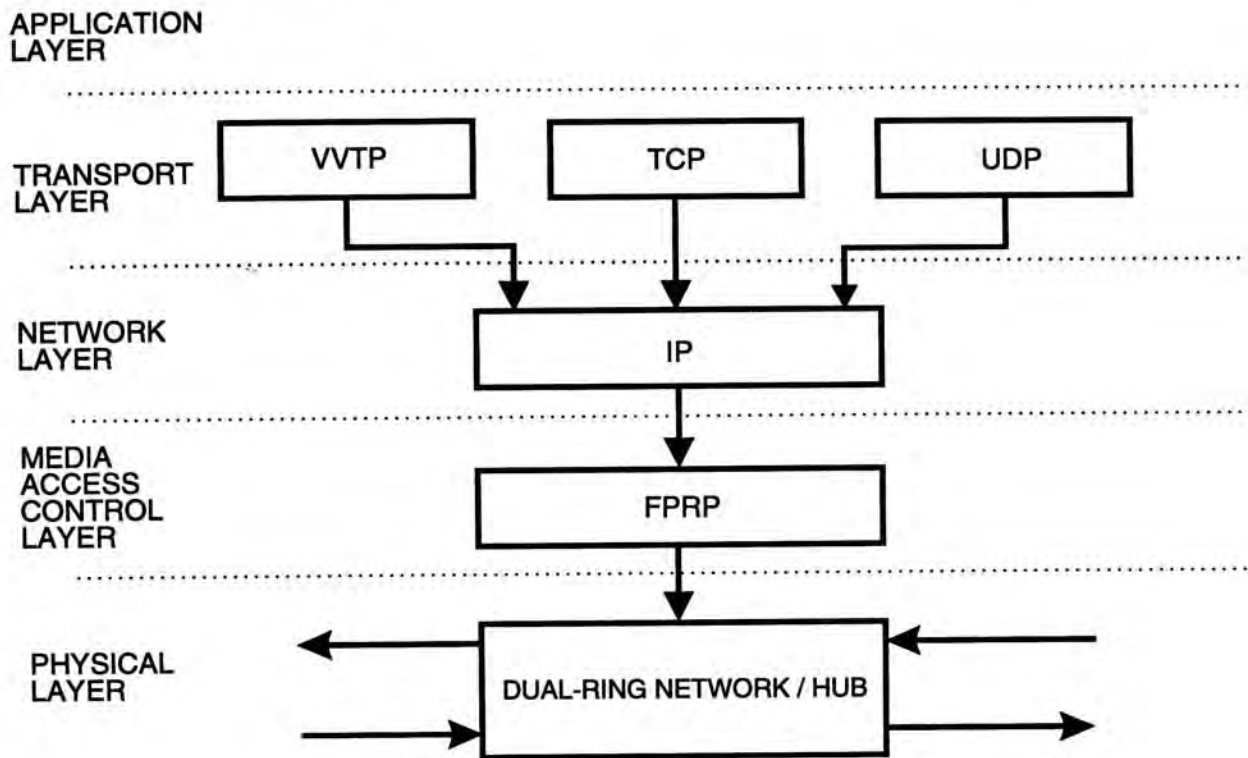


Figure 3.1: Protocol Layering of CUM LAUDE NET

The operating system for CUM LAUDE NET is a public domain system called Linux [22]. The kernel of the operating system has been modified to support VVTP as well as TCP/UDP. VVTP has been given a higher priority than TCP/UDP to prevent non-real-time packets from blocking up the transmission queue.

### 3.3 Segment, Datagram, and Packet Format

Fixed size segments (Transport Layer - 552 bytes), datagrams (Network Layer - 576 bytes), and packets (Fast Packet Routing Layer - 582 bytes) are used in



CUM LAUDE NET. In the FPR layer, a CUM LAUDE packet is consisted of four fields as shown in Figure 3.2:

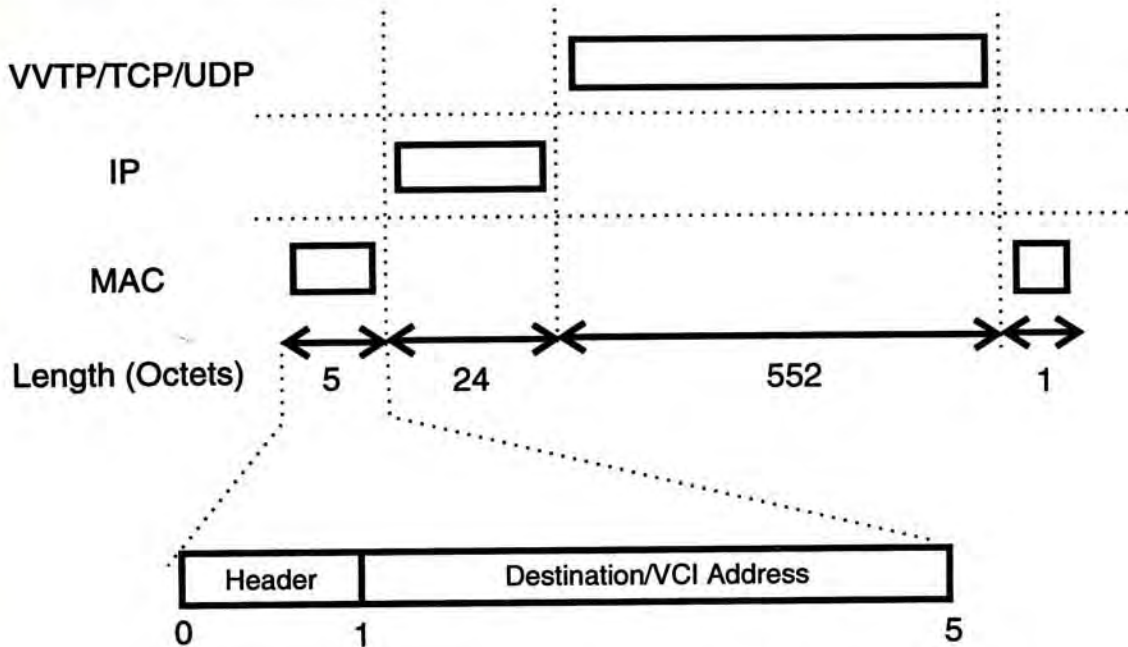


Figure 3.2: Packet Encapsulation of CUM LAUDE NET

1. an access control header (1 byte),
2. a destination address/VCI (4 bytes),
3. a fixed size IP datagram (576 bytes),
4. and an access control trailer (1 byte).

The control information required for fast packet routing is carried by the first 5 bytes (header + destination address/VCI). The access control header and trailer bytes are used for frame synchronization as well as for providing routing information. They will be described in details later.



### **3.3.1 IP/VCI field**

The IP/VCI field in the header (32 bits) is used to identify the address type as well as its destination address. If it is an IP address, the destination address field shows a 32-bit IP address. If it is a VCI address, the destination address field shows a 24-bit VCI. The router/hub will compare the VCI of the incoming packet with a VCI table stored locally to hunt for a match. The VCI is used mainly for multicast applications. It is connection-oriented and will have to be set up during the connection start-up phase. The concept and implementation of VCI are very similar to the ATM standard.

## **3.4 Data Link**

The physical layer of CUM LAUDE NET employs the byte-oriented TAXI data link (Transparent Asynchronous Transmitter/Receiver Interface) [23]. The TAXI data link is a very popular protocol and chip sets with a transmission speed up to 275 MBaud is available at very low cost. The TAXI link has been widely employed by many FDDI, Fiber Channel, ESCON and ATM manufacturers.

### **3.4.1 byte format and data link synchronization**

Each TAXI byte can be an 8-bit, 9-bit, or 10-bit byte, and after encoding, the byte will become 10 bits, 11 bits or 12 bits. The encoding is done by a mBnB block code. We choose a byte format of 9 bits to facilitate the implementation of the physical layer, the media access control layer, the FPR layer and the network layer. In particular, control information and packet boundary can be

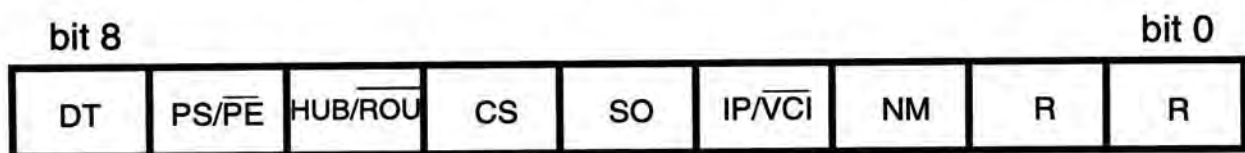
easily detected.

Data bytes are classified into two broad types: the data bytes or the access control bytes. The type is indicated by the MSB of the 9-bit byte. A “zero” indicates that the byte is an access control byte and a “one” indicates that the byte is a data byte. The MSB is called the Data Type field DT (MSB/bit 8). For a data byte the remaining 8 bits simply represents the data. The representation of the access control byte will be shown later.

One of the main reasons for using the TAXI protocol is that the data link is automatically synchronized when the link is powered on. Synchronization is achieved by the transmission and reception of a special SYNC symbol “K28.5”. Synchronization can be achieved with less than eight SYNC symbols according to our experimentation.

### 3.4.2 access control byte

The access control header/trailer byte consists of five fields (Figure 3.3):



DT: Data Type (0: Control, 1: Data)  
 PS: Packet Start  
 PE: Packet End  
 HUB/ROU: Control Byte from Hub or Router  
 CS: Cycle Start

SO: Slot Occupied  
 IP: Internet Protocol Packet  
 VCI: Virtual Circuit Identifier Packet  
 NM: Network Management Packet  
 R: Reserved for Future Use

Figure 3.3: Header format for FPRP (Router)

The interpretations are as follows:

1. The PS/PE field (bit 7) is used to identify whether the current access



control byte is a header or trailer byte. The access control header and trailer bytes are identical in structure.

2. The HUB/ROU field (bit 6) is used by the NIC to identify whether the control byte originates from a router or a hub.

3. If the access control byte originates from the router, then

(a) bit 5 and bit 4 denote the Cycle Start field CS (bit 5) and Slot Occupy field (bit 4) respectively, and are used by the routers to perform the media access control (the ACTA protocol).

(b) The IP/VCI field (bit 3) indicates whether the immediate 4 bytes following the access control header byte is an IP address or a VCI.

(c) The Network Management Packet field NM (bit 2) indicates whether the packet carries network management information. This type of packet is used only by the routers (and not by the hub) for the purpose of network management.

(d) Bit 1 and bit 0 are reserved.

4. If the access control byte originates from the hub, then

(a) The Hub Command field HC [bit5 bit4] are for encoding the hub polling command:

0 0 : Poll node X where X is the polling address

0 1 : Force all node to release the link

1 1 : Allow node X to transmit

The usage of these commands will be described in a later section.



- (b) The Polled Address field PA [bit 3 bit 2 bit 1 and bit 0 (LSB)] represents the polling address which is hardwired to the network interface cards. Since there are 4 bits in the address, up to 16 hosts can be connected to a hub.

### **3.4.3 packet/frame boundary**

The use of the access control bytes in front and after a packet allows a packet to be easily identified. That is, the frame boundary is easily detected and exceptional cases can be taken care of. This simplification of the physical and higher layer implementation justifies the reduction of the bandwidth due to the access control bytes and an extra bit in every byte.

## **3.5 Fast Packet Routing Protocol**

As it has been mentioned above, the CUM LAUDE NET protocols are designed to provide efficient implementation of fast packet routing. The media access and routing algorithms uses only the first 5 header bytes of a FPR packet shown above. That is, when a CUM LAUDE packet is forwarded from router to router or to the hub, the router or hub only reads in the packet header (the first 5 bytes) to determine the appropriate routing action. The IP portion of the packet is not processed by the routers or the hubs at all. Thus, the processing complexities of the multimedia network is greatly reduced. The IP portion will be copied to the hub when an address match is found or will be forwarded with the header to the downstream router.

For Level-2 or Level-1 dual ring hierarchies, the media access/routing control

is based on the ACTA protocol [14] which is described in the next section. For the hubs, media access is based on a polling algorithm.



### **3.5.1 Level-2/Level-1 Bridge/Router**

Fast packet routing algorithms based on ACTA are implemented on all Level-1/Level-2 Bridges/Routers of the dual-ring networks. The data rates for Level-2 and Level-1 nodes are set to be 1-Gb/s and 100-Mb/s on each ring respectively. The dual-ring run in opposite directions.

For incoming packets, the router examines the FPR packet header to determine whether the destination of the packet belongs to a host that is served by the router. If the destination matches, the router will copy the packet to the local buffer for forwarding to the next hierarchy or the local host. The original packet will also be forwarded onwards as shown in Figure 3.4.1. The packet will be erased only at the erasure node according to the ACTA protocol. Thus, it is not necessary for the router to search for information inside the IP datagram.

The details of address comparison is as follows. The destination in the FPR packet header can either be a direct IP address or an indirect address called virtual connection identifier (VCI) which is suitable for multicasting. The two cases are indicated by a single bit in the header. If the destination is an IP address, a direct comparison is made to the address field. If it is a VCI address, the first three bytes in the address field will be used to compare with all VCI addresses in a VCI table dynamically stored in the router. Whenever a match is found, the packet will be routed to the next hierarchy or the local host (Figure 3.4.2). When the packet reaches the local host, the header, address and trailer fields of the FPR packet will be discarded, thus retrieving the original IP datagram.



 Incoming packet  
 Outgoing packet  
 X, Y, Z : Host address  
 E: Empty Packet

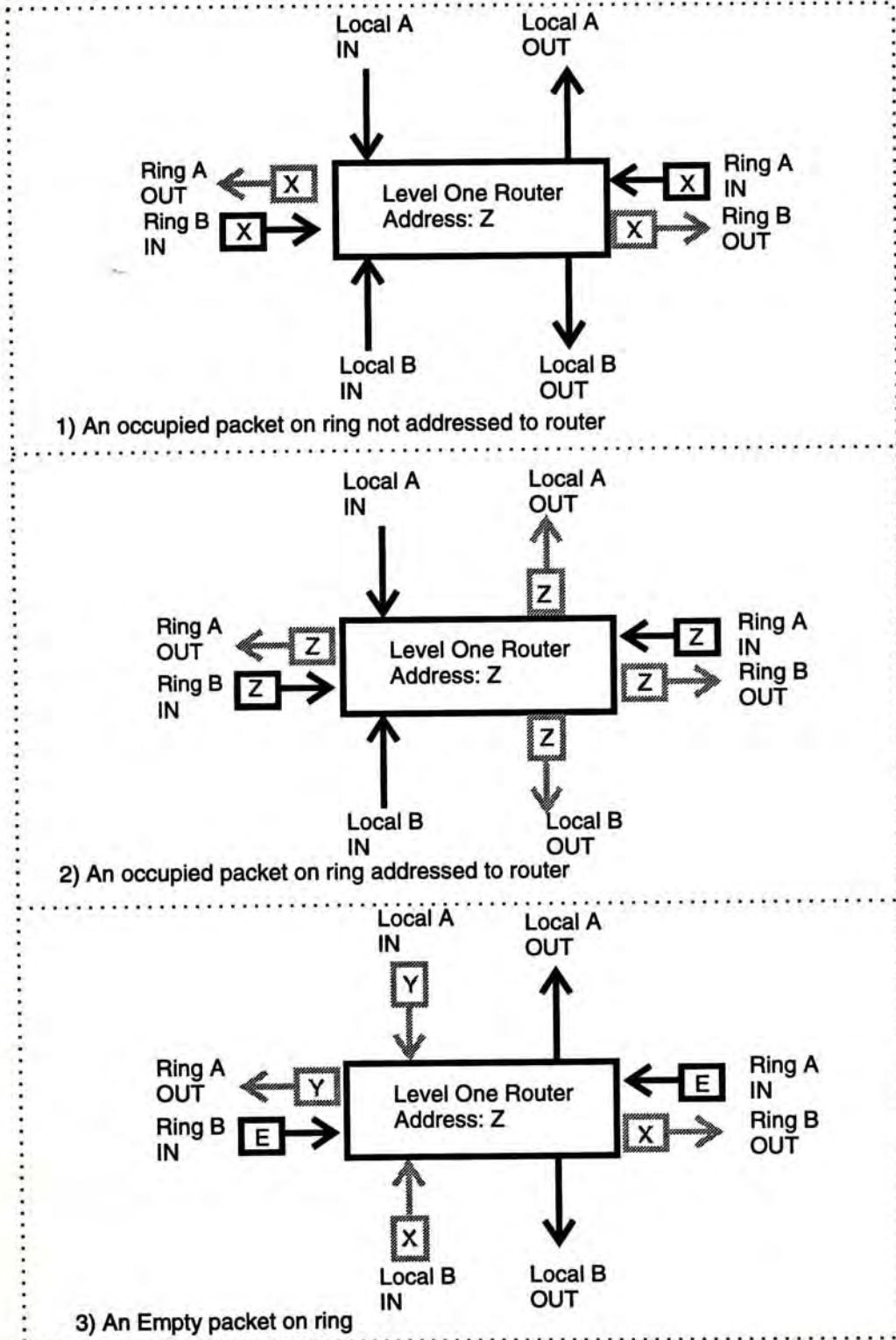
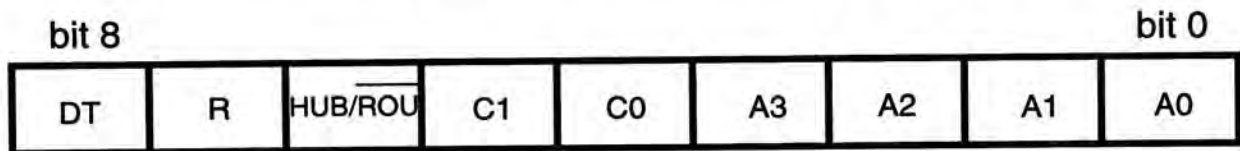


Figure 3.4: Fast Packet Routing

For outgoing packets, the router first determines whether there is any empty incoming packets. When there is an empty slot available, the router can fill up the empty slot with a packet in the queue according to the ACTA protocol (Figure 3.4.3).

### 3.5.2 Level-1 Hub

Each hub connected to a router is used to serve as a concentrator/server to a number of local hosts. Any packets received by the hub will be broadcast to all local hosts connected. The hubs also poll individual local hosts periodically to collect packets that are to be sent into the network. A polling algorithm is used in order to provide orderly packet transmission from local hosts sharing a common broadcast link. The header format is shown in Fig.3.5.



DT: Data Type (0: Control, 1: Data)  
 R: Reserved for Future Use  
 HUB/ROU: Control Byte from Hub or Router

C1, C0: Hub Command  
 A3, A2, A1, A0: Poll Address

Figure 3.5: Header format for FPRP (Hub)

### 3.5.3 Local Host NIC

The local host network interface card (NIC) is directly connected to the host whose function is to process packets going between the network and the host. The NIC interrupts the operating system periodically to make sure that real-time packets can be served timely.



## **3.6 Media Access Control Protocol I : ACTA**

The Adaptive Cycle Tunable-Access (ACTA) protocol is a simple and fair network protocol for loop-bus/ring networks. It is especially suitable for multimedia network because:

- it has a fair network access scheme;
- it has low delay at low traffic load;
- it has high throughput even at high traffic load;
- a simple protocol allows easy protocol implementation;
- its throughput performance is independent of the round-trip delay time, e.g. high network utilization even when only a single node is transmitting;
- it supports multimedia integrated traffic;
- it allows reservation for performance guarantee.

The ACTA protocol is designed for hierarchical dual-ring network. Ring topology has several unique advantages over a centralized switching hub topology. The linear topology allows reserved service guarantee and fair sharing of bandwidth among all nodes. It also reduces the problems due to network congestion and complexity in network control and management. The distributive, sequential arrangement of the nodes also facilitates real-time protocol implementation. Distributive packet routing simplifies packet processing and introduces little packet delay.

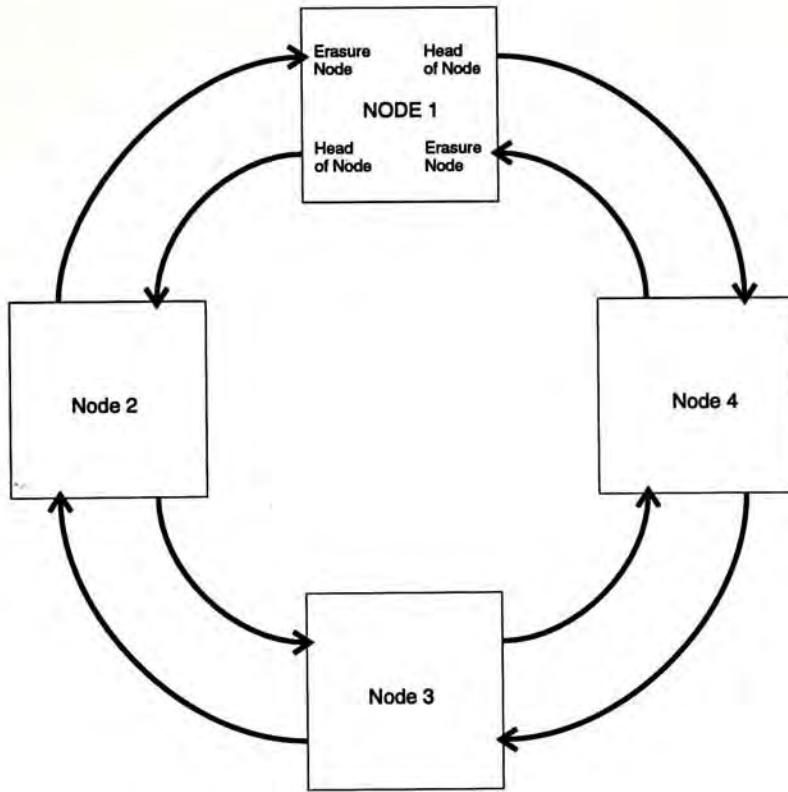


Figure 3.6: The Role of different nodes in a Dual Ring Architecture

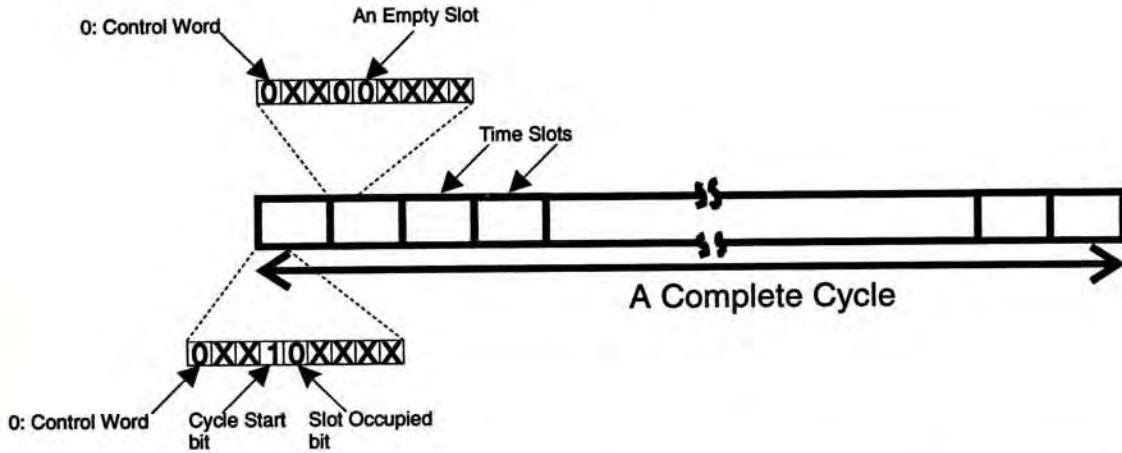


Figure 3.7: A complete cycle of time slots generated by the head of bus

Figure 3.6 shows a physical, 4-node, dual-ring network. Each ring runs in opposite direction to each other. Node 1 performs two functions for each ring. For the outgoing direction, it acts as the head-of-bus. Continuous empty slots are generated by this head-of-bus. The first slot of a cycle has the “Cycle Start” bit



marked. Empty slots are indicated by a clear “Slot Occupied” bit. A complete cycle of time slots transmitted by the head of bus is shown in Figure 3.7.

For the incoming direction, node 1 acts as an erasure node to remove all occupied or empty slots from the ring. The number of occupied slots within a cycle is used for predicting the traffic loading and for adjusting the number of time slots on next cycle. This is an important characteristic of the ACTA protocol. The operation of ACTA protocol scheme is as follows:

1. Fixed-size empty slots are generated continuously from the head-of-bus node to the two opposite direction rings as shown in figure 3.8. Each slot has two control bit: Cycle Start and Slot Occupied. The Cycle Start bit is enabled only by the first slots of a cycle. It is disabled for the rest of slots and will not be modified by any other node. The slot occupied bit is disabled when it is first generated by the head of bus. It is then enabled when the empty slots are occupied with data packets.

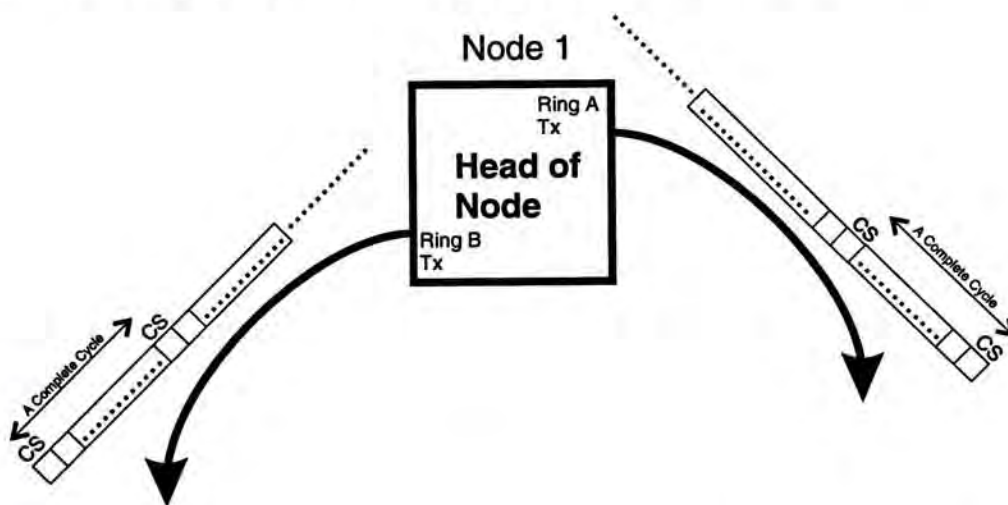


Figure 3.8: A cycle of empty slot generated by the head of bus

2. The number of slots within a cycle is set by a cycle-length counter kept by the head of bus.

3. Cycle Start bit can only be enabled by the head-of-bus during the first slot of a cycle.
4. Any intermediate node (node 2, 3, 4) on receiving the empty slots, can write consecutively onto the first  $N_q$  (the assigned quota) empty slots after a cycle-start bit is detected. The quota,  $N_q$ , is set in advance. The number is varied on different node according to the priority. Nodes with heavy traffic load will assign a larger  $N_q$ . After a node write a packet onto the empty slot, the slot occupied bit is enabled.
5. There are two conditions that a node must stop transmission and wait for the next cycle start.
  - (a) A node has used up all the quota ( $N_q$  slots) in that cycle;
  - (b) A node has nothing to send without use up all its quota.

The quota is reset to  $N_q$  in both cases and wait for the start of next cycle.

6. For the erasure node, all the occupied slots or empty slots is removed. The number of occupied slots is counted within a cycle. It is used to predict the traffic loading and is used to calculate the cycle length of next cycle. The new cycle length is kept at the cycle length counter.
7. After the current cycle has been completed, the head-of-bus node initiates a new cycle start with a cycle length given by the cycle-length counter.

The media access scheme is very simple. Each node only need to detect the Cycle Start and Slot Occupied bits of each slot before it can access  $N_q$  consecutive slots within a cycle. This requires only a few tens of program instructions



to implement the protocol. The performance of the ACTA protocol has been shown to be very good [14].

## **3.7 Media Access Control Protocol II: Hub Polling**

The hub polling protocol is used to avoid contentions among the local hosts attached to the same hub. As has been explained before, each hub may serve up to 16 host stations because a single host can rarely use up the bandwidth available to each hub. By avoiding packet collisions, hub polling allows real-time services to be scheduled easily. The protocol was designed by another member of the CUM LAUDE NET team (Ringo Lam).

The hub polling control information is contained in the access control byte. As described before, the HUB/ROU bit in the access control byte is used by the network interface card on each host to distinguish the hub polling commands from the router's access control commands.

The format of the hub polling command is shown in Figure 3.5. The 4-bit Polled Address field allows the hub to poll up to 16 hosts (network interface cards). There are three hub commands packed in two bits, namely, poll node X, force all node release and allow node X to transmit.

For the downlink (from the hub to the local hosts), the hub simply broadcasts all packets received from the router to all hosts. The synchronization for the downlink is maintained all the time since the hub broadcasts to the attached hosts constantly. Hub polling commands from the hub are inserted into the downlink from time to time.

For the uplink (from the local hosts to the hub), the hub polling protocol is as follows:

1. Before any polling, the transmission link from the NIC to the hub is assumed to be in the idle state. The receiver of the HUB are out of synchronization.
2. The hub broadcasts a "Poll Node X" command together with polling address in a single access control byte. This polling command can be inserted into any position of a packet being broadcasted by the hub to some host. This is essential to avoid too much delay in the polling.
3. The command together with the node address will be decoded by the appropriate host or network interface card. The transmitter of this host will be turned on when the polling address matched with its own.
4. After the transmitter has been turned on, a SYNC signal "K28.5" will be transmitted. The sync signal will cause the receiver at the hub to get in synchronization with the local host. The hub then broadcasts an Allow Node X command to the downlink to grant permission to the local host to send.
5. After the local host (network interface card) has received this Allow Node X command, it will start up its state machine to transmit data out automatically until it encounters the end of packet, at which point it will send an access control trailer byte to signal the completion to the hub.
6. On receiving the access control trailer, the receiver at the hub will transmit a Force Release All command to finish the polling sequence.



7. The hub then proceeds to poll a second host.

### 3.8 Protocol Implementation on CUM LAUDE NET

The ACTA protocol has been implemented on the CUM LAUDE NET. In order to facilitate the implementation, the protocol has been slightly modified without changing the spirit of the original ACTA protocol.

In the original ACTA protocol, fixed-size empty time slots are generated consecutively by the head-of-bus node with the Cycle Start bit of the first time slot marked. All the other time slots are generated with a clear Slot Occupied bit. In the implementation of the ACTA protocol, it is found that the generation, examination and removal of each arriving empty slot at each router consume a lot of processing time.

In order to reduce the amount of processing, we discovered that it would not be necessary to generate or transmit the empty time slots from the head-of-bus because our physical links are automatically synchronized by the TAXI protocol. Instead, a small control packet is generated and transmitted at the beginning of each cycle with a cycle length proportional to the number of empty slots available. The control packet represents the start of a cycle, and contains information about the number of unreserved empty slots (two bytes) and reserved empty slots (two bytes) available in that cycle.

The number of priority levels can therefore be easily extended by appending a two-byte integer  $N_i$  at the end of the control packet for each priority level. Each integer  $N_i$  simply represents the number of available slots for each priority

level in that particular cycle. In our present implementation, only two priority levels are used and the control packet has a total length of 6 bytes, including the access control header and trailer bytes.

The format of the control packet is shown in Figure 3.9. An active Cycle Start bit indicates that the packet is a control packet. When a router receives the control packet, it holds the control packet in its buffer. If there are still empty slots available in that cycle as is indicated by a positive integer  $N_i$ , where the subscript  $i$  denotes the priority level, the router can append its data packets immediately after the previous data packets up to a total of  $N_q$  provided  $N_q \leq N_i$ , where  $N_q$  is the assigned quota per cycle.

After the slots are written, the control packet is released with  $N_i$  replaced by  $N_i - N_q$ . If  $N_i \leq N_q$ , only  $N_i$  data packets can be sent by the router. This is because  $N_i$  must be a positive integer, meaning that the number of packets that can be transmitted by any node in a particular cycle cannot be greater than the number of empty slots still available. The operation is repeated for each priority level before the control packet is released.  $N_q$  and  $N_i$  limit the number of packets that can be transmitted by each node. The number of reserved slots at different priority levels (the initial  $N_i$ ) represents different bandwidths allocated to different priorities of traffic.

In this way, the generation and transmission of empty slots are substituted by a simple integer in the control packet. Such an implementation of the ACTA protocol is very similar in spirit to the implementation of FDDI. The similarities and differences will be further elaborated in a future publication.

To summarize, the ACTA protocol implementation is as follows:



1. The head-of-bus node generates a new cycle with a control packet containing information about the cycle length and the number of empty slots available for each priority.
2. After receiving the control packet, each router can transmit their data packets up to their preassigned quota for each priority level.
3. The number of available empty slot field in the control packet is decremented by the number of data packet that has been transmitted for each priority. The control packet is then released.

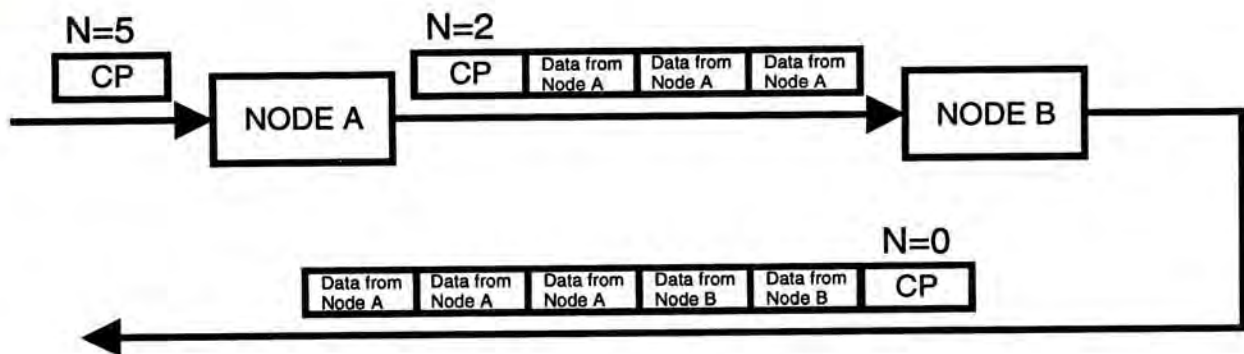


Figure 3.9: Packet transmission by Node A and Node B

As an illustration (Figure 3.9), a control packet (CP) and a new cycle is generated by the head-of-bus node. Each node can append its data packet to the cycle immediately after receiving the control packet. For simplicity only one priority is shown and the number available empty slots is  $N = 5$ . For node A, after the control packet has been received, three data packets are transmitted. The control packet is then appended at the end of the three data packets. The number of available empty slot field  $N$  in the control packet is decremented by three to  $N = 2$ .

The same scenario occurred for node B. Since there are three data packets in front of the control packet, Node B will simply forward these three data packets before receiving the control packet. Since  $N = 2$ , node B can only transmit 2 data packets in this cycle even though it has 3 packets in its queueing buffer. The control packet is then appended and the number of available empty slots  $N$  is now zero. All the other nodes must wait for a new cycle (with a non-zero  $N$ ) before they can transmit.

The present implementation have some tradeoffs. First, the present implementation can only be achieved by an active regenerative link. Second, each node can only transmit once during each cycle whereas in the original ACTA protocol, each node can send its data at any time within a cycle. This means the current implementation will have a slightly bigger delay jitter. Overall, these two tradeoffs are justified with respect to the simplicity gained in the protocol implementation.



## **Chapter 4**

# **Hardware Implementation & Performance of Routers and NIC**

In this chapter, the actual implementation of the routers and NICs will be described. The motivations for this particular implementation has been described in Chapter 3 and will not be repeated. Three generations of routers and NICs has been tried, with the later generations being improvements of previous generations. The performance of various generations of routers and NICs and the experience learned will also be described.

### **4.1 Functionality of Router**

A block diagram of the router is shown in Figure 4.1. The router controls packet access to the network. Each router serves two rings running in opposite

directions. The router employs fast packet routing and the ACTA protocol is implemented on the MAC layer.

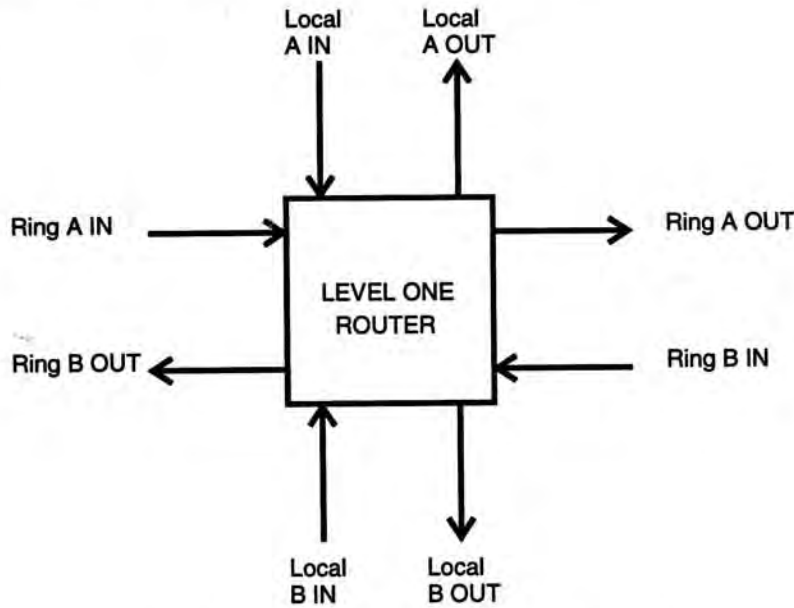


Figure 4.1: IN OUT path of Level One Router

Thus, each router has two pairs of TX / RX modules (one pair for each ring). There are two input and two output data queues, one for each ring. The input queues are used to buffer packets sent from the hub / host stations to the corresponding ring. The output queues are used to store up any arriving packets that should be routed to the local hosts through the hub.

Packets received by router will fall into three possible categories:

1. Data packets not addressed to the local host stations:



Figure 4.2:



Action: Ring A IN to Ring A OUT

Ring B IN to Ring B OUT

2. Data packets addressed to the local host stations:

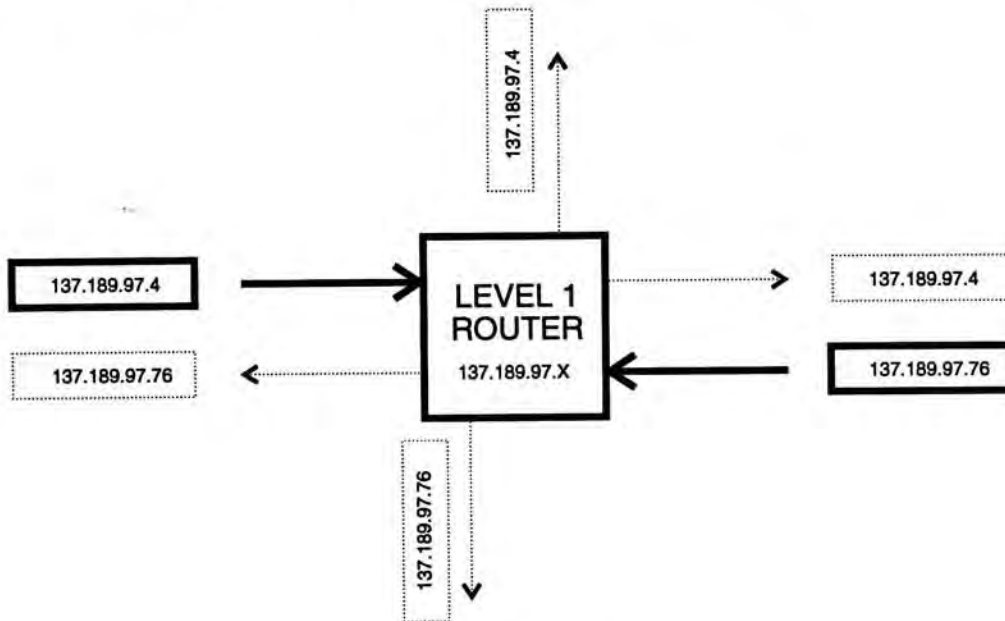


Figure 4.3:

Action: Ring A IN to Ring A OUT & Local A OUT

Ring B IN to Ring B OUT & Local B OUT

3. Access Control Packets (Cycle Start Reset):

Action: The access control packet is firstly removed from the network. The number of empty slots available  $N_i$  is identified from the control packet. Each router can sent at most  $N_q$  data packets to the network. The number of empty slots in the control packet is modified to  $(N_i - N_q)$  and the control packet is released after the data packets have been transmitted.

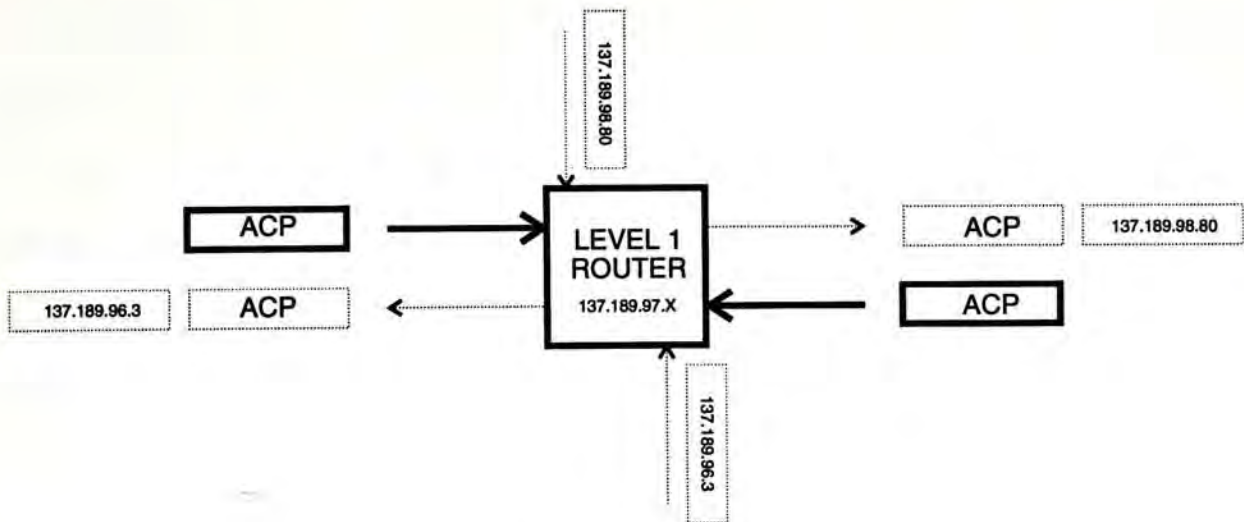


Figure 4.4:

For packet with multicast address, if the destination nodes appear in both the upstream and downstream directions, the same packets will be sent out to both ring A and ring B.

## 4.2 Important Components Used in the Router Design

Before the discussion of the router design, two main components necessary for the router design: the TAXIchip and FIFO memory will be described first. Both components play an important role in the router design:

### 4.2.1 TAXI Transmitter and Receiver

The TAXI transmitter and receiver is a general purpose interface for very high-speed (40 - 125 Mbaud per second serially) point-to-point communications over coaxial or fiber-optic cables. The TAXI chipset emulates a parallel port. Parallel



data is transmitted by TAXI transmitter and is received by the TAXI receiver. Between the two ends, a long serial transmission link is used.

The TAXI transmitter encodes the parallel 8-bit, 9-bit or 10-bit data into 4B5B or 5B6B format according to the parallel data width. The parallel data is then time-division multiplexed into a high-speed serial bitstream and is transmitted in pseudo ECL format.

The TAXI receiver receives the serial pseudo ECL bit stream, demultiplexes the serial stream back to the parallel format, and then decodes the 4B5B or 5B6B format back to the 8-bit, 9-bit or 10-bit parallel data. The TAXI receiver also provides violation signal to indicate any receiving error (such as lost of synchronization) during transmission. A sync signal is provided to the receiving host to indicate that synchronization has been achieved (from the transmitting end to the receiving end).

### **4.2.2 First-In-First-Out Memory (FIFO)**

The FIFO memory [24] consists of two sets of data bus and control bus. The data bus is 9-bit wide. One set of data bus acts as an input port with a write pin to allow data to be written in. Another set of data bus act as an output port to allow data to be read out. Reading and writing are done independently and the FIFO has an internal locking mechanism to resolve conflicts.

All the addressing is done internally. There are two address pointers inside. The starting address pointer points at the start of the data queue while the ending address pointer points at the end of the data queue. A “read” or “write” signal to the FIFO causes the starting or ending address pointer to increment by 1 respectively. When the distance between the starting address pointer and the

ending address pointer is zero (an overlapping of pointers), an empty flag signal is enabled to indicate that the FIFO is empty. When the distance between the two pointers is over half the size of FIFO, a half-full flag is enabled. When the distance between the pointers reaches the full-size of the FIFO, both the half-full and the full flags are enabled.

### 4.3 Design of Router

There are three generations of the level-1 router design. The performance of the third version (latest) is the best. The sections below will discuss the design of the different versions.

#### 4.3.1 Version 1

The version 1 router is the simplest among the three versions. The router is a single PCB board plugged into a host PC. The router does not have any processor and it is completely controlled by the host PC. For this version, the packet transfer rate is limited by the speed of PC and its bus bandwidth and the network throughput is very low.

#### Hardware Design

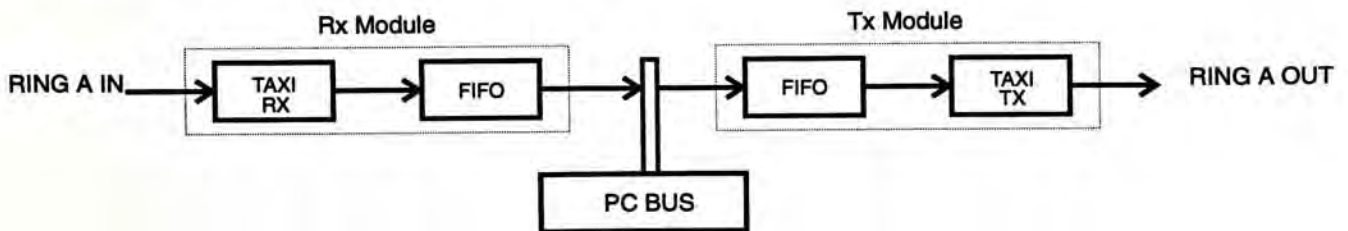


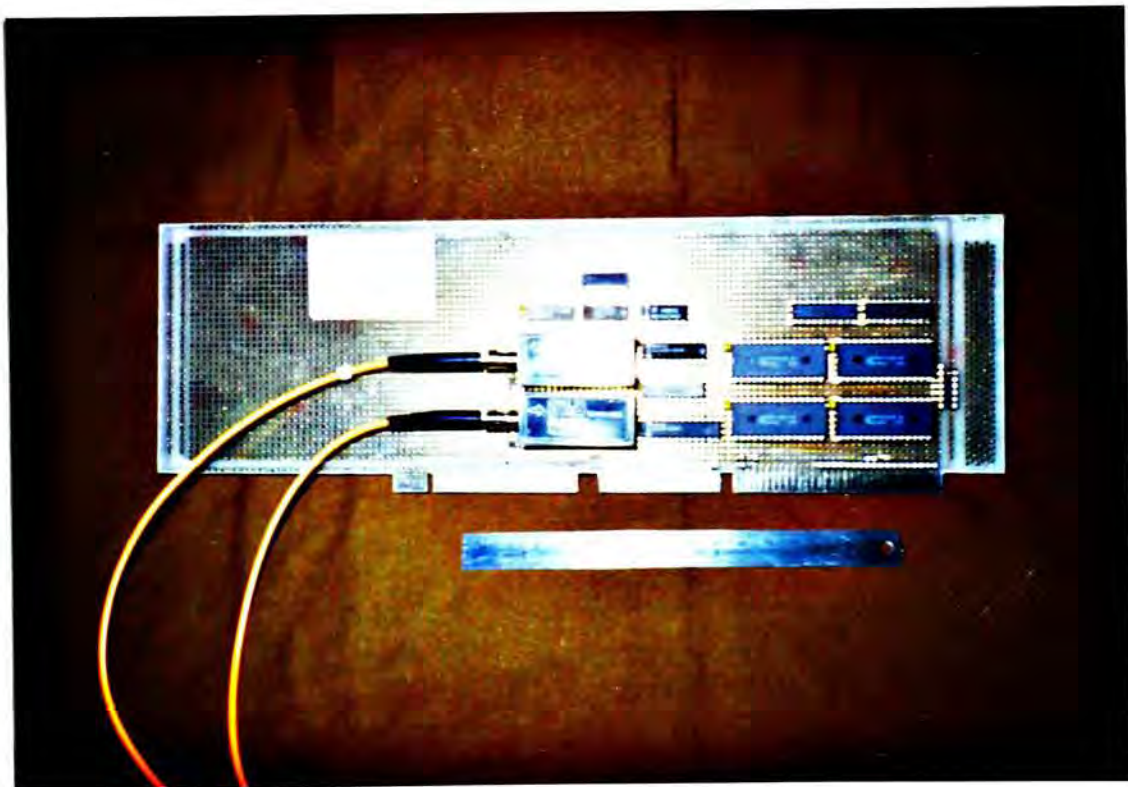
Figure 4.5: Design of Version one Router



Figure 4.5 shows the design of the version-1 router. Only a single ring was implemented for testing purposes. The router consists of a pair of TAXI transmitter and receiver running at 100-Mb/s. Each TAXI is equipped with a FIFO memory to buffer any data from the host or from the network.

There is one status register (not shown in the Figure) that can be accessed by the PC. The status register contains the status of the three flags (empty, half-full and full flag) of the FIFO memory for both the transmitter and receiver. Thus the PC can receive the status information about a complete packet arrival and to read and process the packet accordingly.

A performance test has been performed by connecting three version-1 routers in a single ring. A driver has been installed on each host machine. The driver is a small program routine to handle all the service requests to the router like status read, packet read and packet write.



## **Performance**

Because of the simple design, the performance of the router is not quite satisfactory. The test result of FTP throughput from one router to another is found to be around 4 Mbps. There are two reasons for the low data transfer rate. First, there is a memory bandwidth bottleneck at the PC/AT-bus. Although memory mapped technique has already been used, the maximum PC/AT-bus throughput is found to be 2 Mbyte per second. This speed is the maximum possible data transfer rate, assuming that the entire PC is performing data transfer only and not any other task. The reading and writing done independently by the PC further halves the net throughput. Second, since the PC serves as the router controller, there is also many other task that the PC must run in the background which further lowers the network throughput. The measured throughput seems to confirm these limitations.

### **4.3.2 Version 2**

Version 2 is a big advance over the version-1 router. This version aims at resolving the two major problems caused by version-1 router mentioned above: the PC bus speed bottleneck and the packet processing and control.

#### **Hardware Design**

Figure 4.6 shows the design of version 2 router. It is still a plugged-in single board design similar to that of version 1. First, to increase the I/O throughput of the PC, the VESA local bus (VL-Bus) [25] is used. The VESA local bus is 32 bit wide and runs at the PC's processing speed. From the investigation, it was



found that the PC is capable of transferring data at 100Mb/s over the VL-Bus. Thus, the bit rate of the TAXI transmitter and receiver can achieve 100 Mb/s in this version,

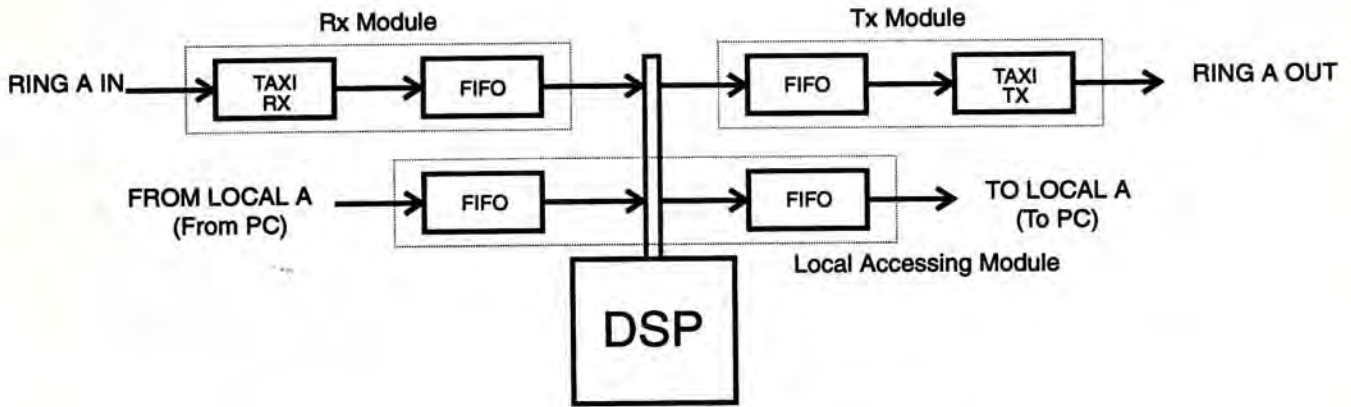


Figure 4.6: Design of Version Two Router

Second, to reduce the work load of the PC, an outboard processor can be used in the router [26]. The advantage of this approach is that the host PC is relieved from most of the tedious work in handling the packets, but the complexity of the router is considerably bigger. The processor must handle all the packet forwarding as well as address recognition. Data is actively moved by the processor from the receiver FIFO to the transmitter FIFO for every packet. Since the processor is actively involved in the data transfer, a very high speed processor is necessary.

The DSP ADSP-21020 [27] has been chosen for this purposes. This DSP has two set of memory buses. The program memory bus is for executing the program code whereas the data memory bus is for transferring data. Due to this characteristics, the DSP can do program execution and data transfer at the same time. Other generic processors mix the two buses and therefore cannot execute program and transfer data simultaneously, thus have a lower bus bandwidth and

efficiency.



### **Performance**

The version-2 router design resolves the two difficult problems of the version-1 router. However, the internal data transfer of the router becomes another bottleneck. The data transfer rate for our version-2 router was tested to be around 20 Mbps. This is because many clock cycles were consumed in performing packet transfer. From this we learn that there is a need to prevent the processor from actively handling the data packets. The solution to this problem is incorporated in the version-3 router design.



### 4.3.3 Version 3

#### overview of hardware design

Figure 4.7 is a block diagram showing the design of the version-3 router. This version resolves all the problems of the previous two versions. We have also implemented two rings in this design. The design is symmetrical with respect to the two rings. The core parts are the two transceiver modules (for Ring A and Ring B respectively), the two local accessing modules (Ring A and Ring B), and the ring controller module.

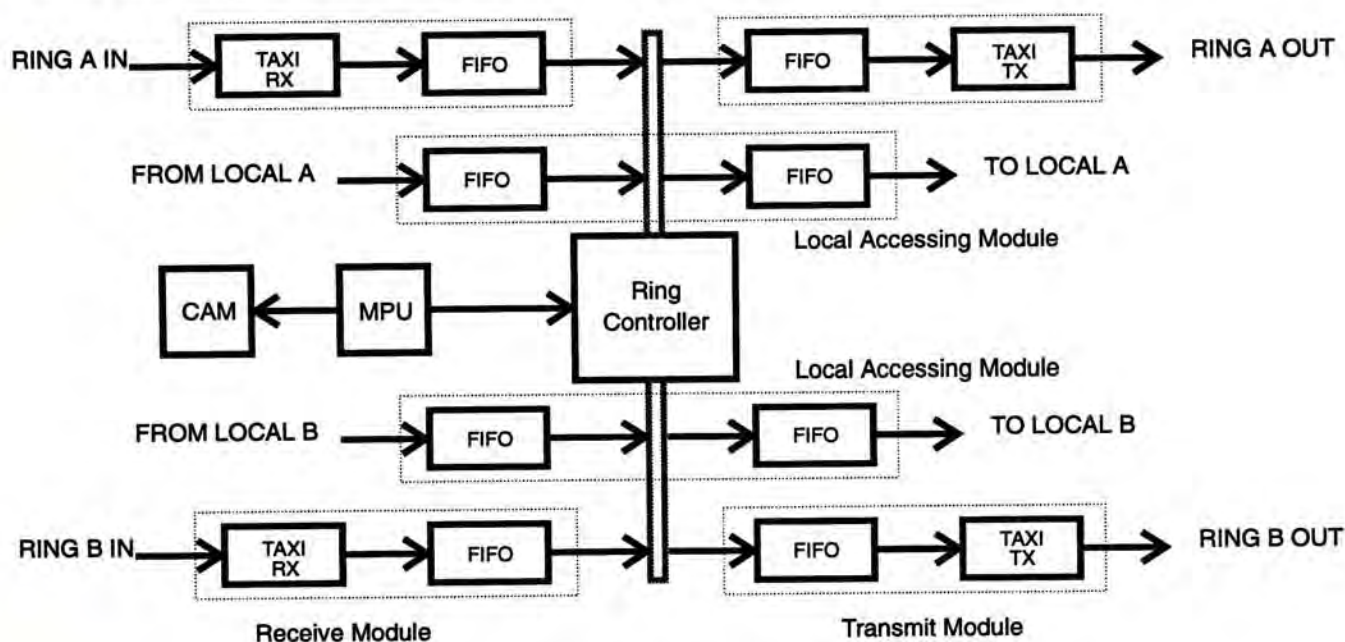


Figure 4.7: Design of Version Three Router

Each transceiver module consists of a TAXI transmitter, a TAXI receiver and some FIFOs, and is very similar to the entire dumb router of the previous versions. The transceiver modules are directly connected to the local accessing modules, which are responsible for forwarding/receiving packets to/from the local hub and host stations.

The ring controller module consists of a microprocessor (MPU), a content

addressible memory (CAM) and some control logic (GAL). This module is responsible for controlling the operation of both rings. First, it controls the internal high speed data transfer between the transceiver module and the local module. Second, it provides a means for the MPU to read from or write to the transceiver modules or the local accessing modules. This is essential for testing and control purposes. Third, it controls how the data paths are connected. That is, the ring controller directs the DMA transfers among various modules (both transceiver modules and local access modules) in which the MPU is not involved in the actual data transfer.

The router circuit has been prototyped, tested and then put on PCB boards. Since the entire router circuit is too large to be place on a standard size PC card, the router is implemented by two PCB boards. One board is for the Ring Controller Module which is equipped with a DSP. The other board contains the transceiver module and the local accessing module.

**normal routing of packets**

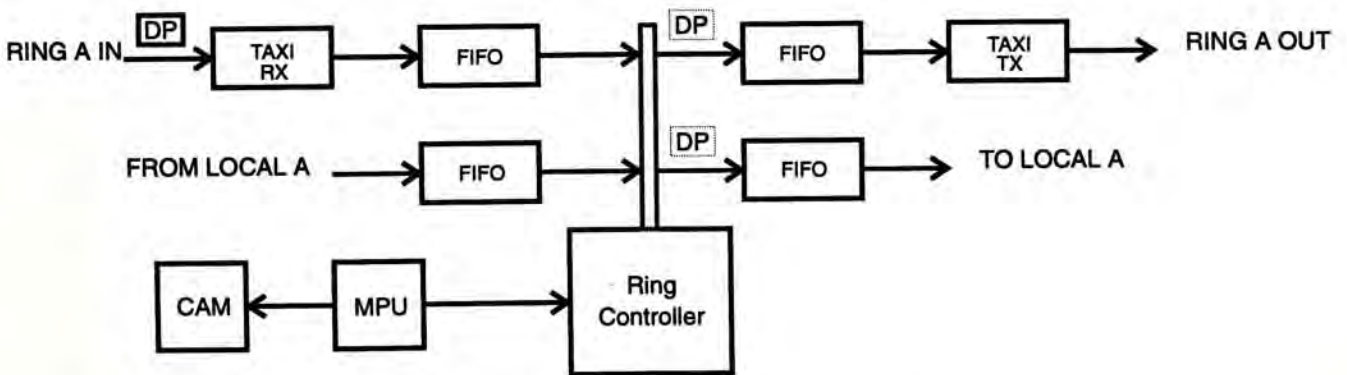


Figure 4.8: Internal Data Transfer of Router

Figure 4.8 shows how a data packet is routed from a transceiver to a local host. The header of the data packet is first read by the MPU in the ring controller.



This is done by first enabling the data path from the transceiver to the MPU and then transferring the header of the data packet in the FIFO of the transceiver module to the MPU. The address field of the packet header is determined.

If an address match is found by the MPU as shown in Figure 4.8, the data path to both the transmitter FIFO (in the transceiver module) and the receiver FIFO (in the local accessing module) are enabled. The header is then written to both FIFO simultaneously by the MPU. The MPU then enables the data path from the TAXI receiver FIFO to the TAXI transmitter FIFO and the receiver FIFO of the local accessing module. Direct memory transfer is then initiated by the MPU to move the remaining packet from the TAXI receiver FIFO to the TAXI transmitter FIFO and to the local accessing module. The data packet is copied to both the local accessing module and the TAXI transmitter module as is required by the ACTA protocol implementation.

The MPU can only control the start of the direct memory transfer. The end of memory transfer is controlled by a control word (access control trailer byte) encapsulating the end of the data packet as described in previous sections. After the DMA has been started, the MPU is free to serve the other ring. The Ring Controller has been designed in such a way that both rings operate independently. While the Ring Controller is performing direct memory transfer on one ring, it can also initiate direct memory transfer on the other ring.

### **fault tolerance routing of packets**

The advantage of a dual-ring network is to provide single fault protection for link or node failures. A dual-ring network can handle these kinds of faults. Figure 4.9 shows the case of link failure between Node 2 and Node 3. The data paths

inside Node 2 and Node 3 must be rerouted to avoid the faulty link.

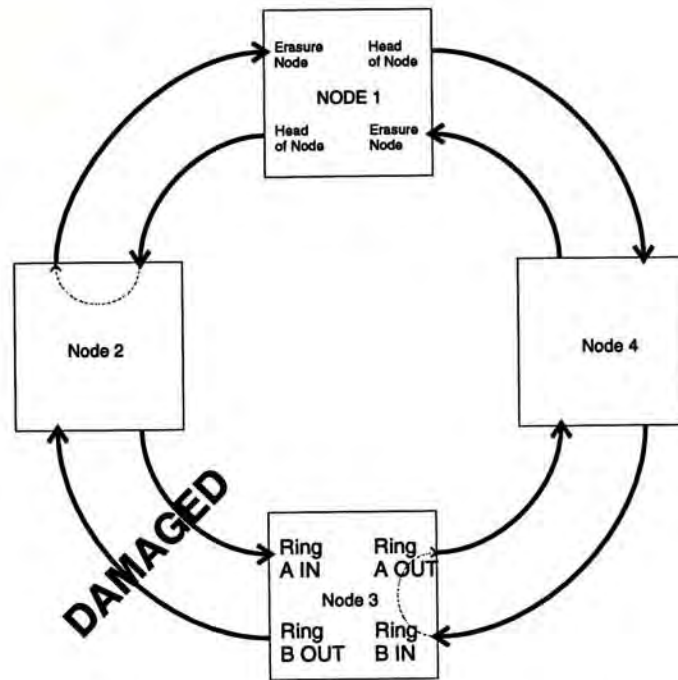


Figure 4.9: Scenario of Link Reconfiguration on Link Fault

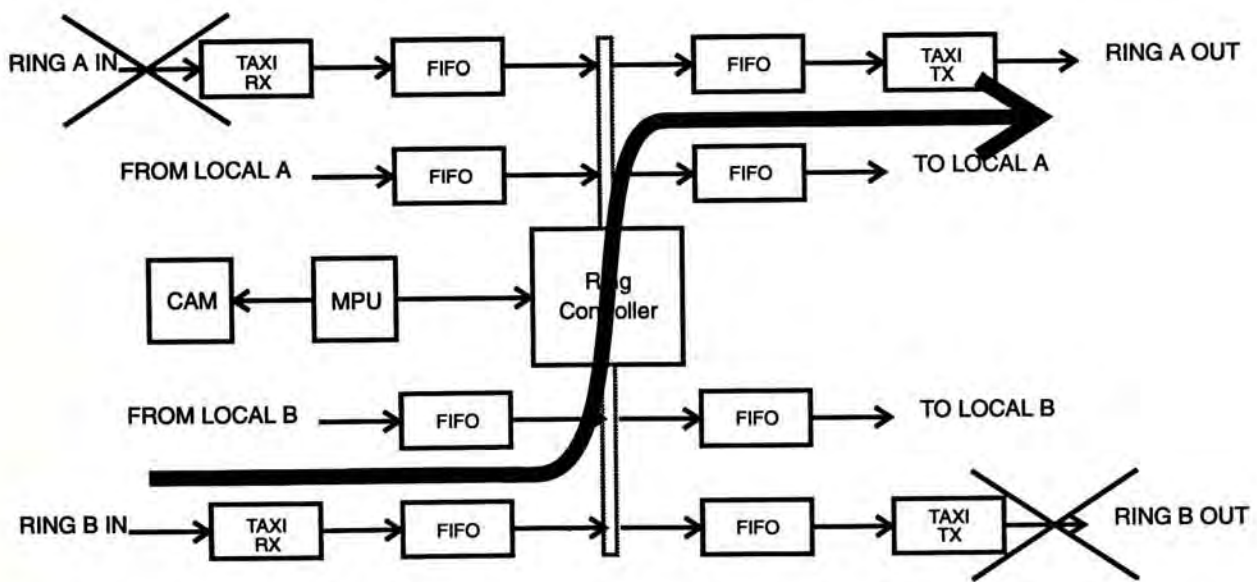


Figure 4.10: Internal Data Transfer of Node 3 from Ring B to Ring A on Link Fault

Figure 4.10 shows how the data paths within a node can be reconfigured to avoid the faulty links. All data packets, instead of taking the normal paths from



A IN to A OUT and from B IN to B OUT, will take the path from B IN to A OUT. In this way, single-ring operation and full-connectivity are maintained, demonstrating the fault tolerant characteristics of the dual-ring network.

The Ring Controller is implemented by advanced programmable logic technology so as to reduce the complexity and board spaces of the router. The total time delay of the gate array logic is also guaranteed.

### **multicasting**

In packet multicasting, a virtual address must be used by a host station for transmitting packets from a single source to multiple destination stations. This virtual address can be considered as a virtual channel identifier. It is allocated by the network manager. Each destination host station and all intermediate network routers must store this virtual multicast address in a virtual address table that is maintained locally and dynamically when connections are set up. Any packet with a multicast destination address will be collected by the appropriate destination nodes.

The number of multicast addresses must be limited. First the storage of addresses consumes memory and second, the searching of multicast addresses is computationally intensive. If too many multicast addresses are kept by a node, the search may require too much CPU time which may slow down the routing of other packets.

A Content Addressable Memory (CAM) [24] is used for this purpose. The CAM is a 256 byte deep 48 bit wide memory. Each 48-bit memory is a register for holding the address. There are at most 256 addresses that can be kept in the CAM. The 256 registers act as a parallel comparator. Once the CAM is set



to the searching mode, any data appear on its data bus will be used to compare with all its registers' content simultaneously. Since the comparison is done in parallel, there is no need to perform searching sequentially. The CAM needs only 70ns to determine whether an address is in the table. The limited size of the CAM also means that the number of virtual connections per node is limited.



### **Performance**

A test has been set up to measure the router throughput (in terms of packets per second). The configuration of the test is shown in Figure 4.11. To reduce the testing complexity, a loopback test is chosen. That is, the transmitter is connected to the receiver on the same ring. To prevent the bottleneck of data transfer due to the bus bandwidth and processing speed of the PC, the PC is not involved during the testing.

A packet is put into the transmitter FIFO by the DSP in advance. The



packet is an IP packet with an address matching that of the local host. When the transmitter is enabled, the packet in the Tx FIFO is transmitted out. The packet is received by the TAXI receiver and is put into the Rx FIFO. The packet header in the receiver is processed by the DSP. Since the packet belongs to the local host, the packet will be directed to both local FIFO and the TAXI TX FIFO as described before. A similar process is performed on ring B. The router therefore sees a packet continuously appearing on both rings with every packet all addressed to the local host. This test gives the peak transfer rate of the router. A timer is kept inside the DSP to count the number of packets being routed by both rings within a certain period of time.

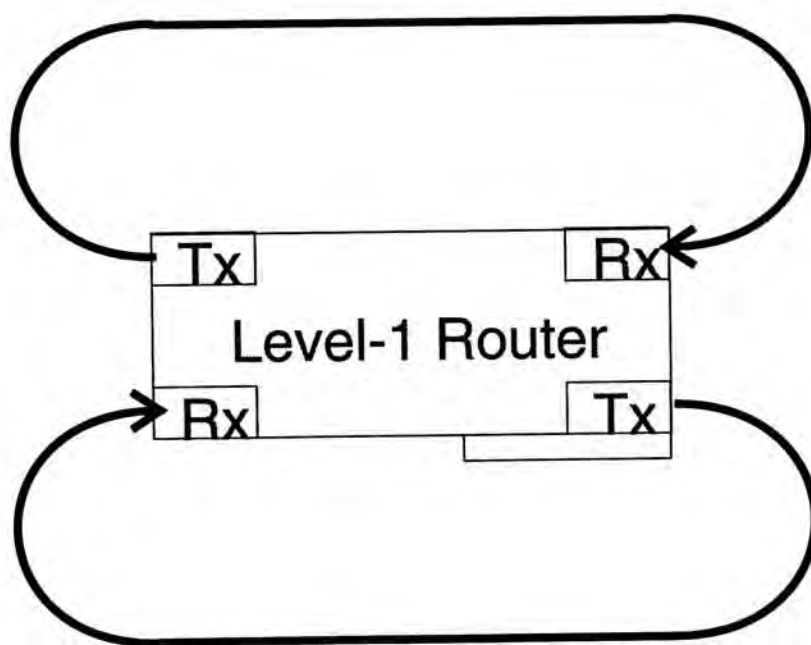


Figure 4.11: Configuration of performance test on Version 3 router

The result shows that the version-3 router performs better than the previous two versions, achieving a full network speed of 100 Mb/s. Since the testing avoid the PC bus bottleneck, the router can perform full speed packet routing.

## **4.4 Lessons Learned from the High Speed Router Design**

We have learned many valuable lessons from our experimental prototyping efforts. It is only through the prototyping effort (especially from the three generations of network router designs) that we understand many key issues and problems. Careful considerations are needed in the design of high-speed integrated networks. Our experiences can also be applied to many other high-speed networks designs, such as FDDI and ATM. Below are a list of the lessons learned.

1. Components should be carefully selected. Inappropriate component selection will not only increase both the chip count and cost but also affect the timing tolerance of the entire circuit which is very difficult to trace. The components used should be programmable during the prototyping stage so that any change in the circuit design or logic functions can be easily accomplished.
2. A well-balanced work load between the router and the hosts should be clearly defined. This will increase the efficiency of the whole network.
3. Some flexibility of protocol implementation should be allowed. The network interface unit should be designed in such a way that any protocol change will not cause any hardware modification. This is particularly important during the development period.
4. Critical functional parts should be hardware to prevent the active involvement of the control processors in the router or the NIC. This is because

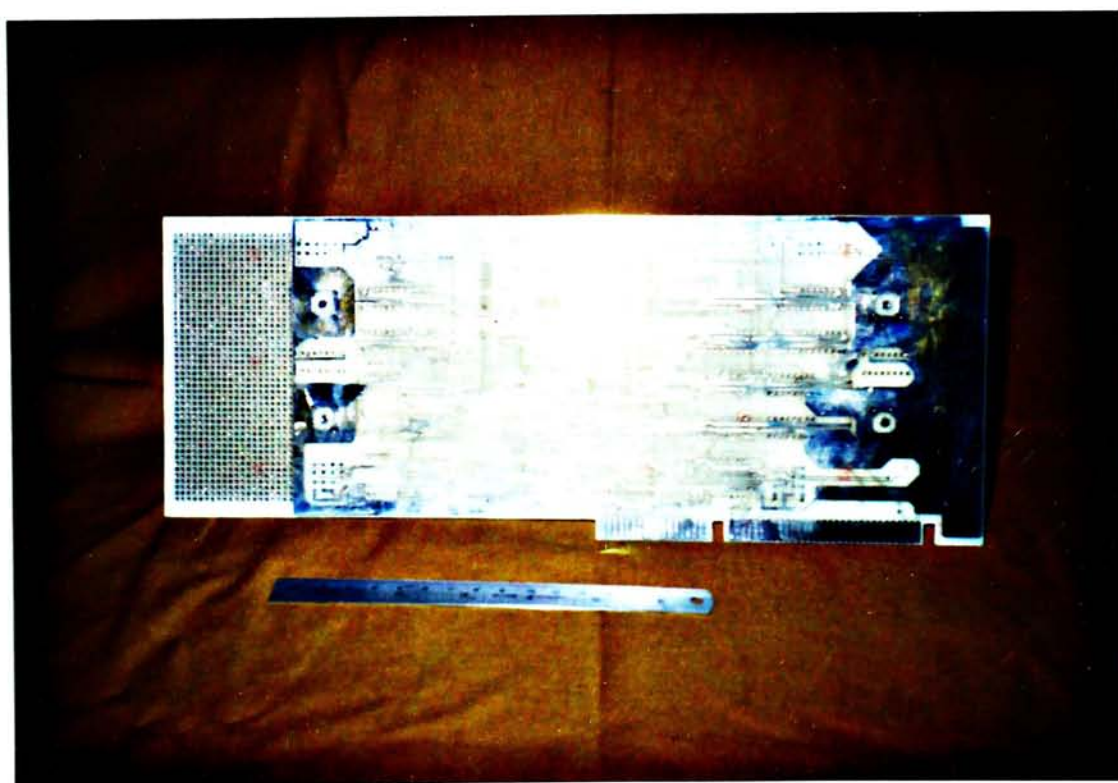


the use of processors will limit the overall throughput. Like the router design in the final version, high speed data movement is done by DMA automatically instead of by programmed IO involving the DSP.

PCB Layout for DSP Controller

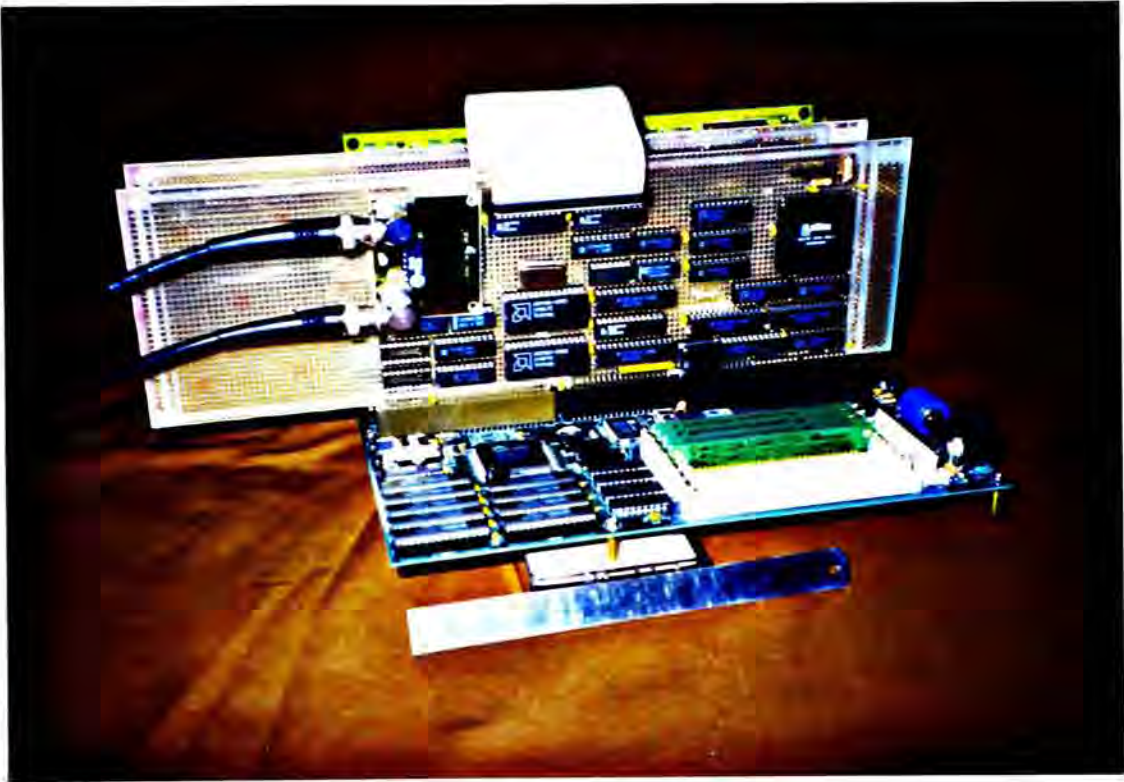


PCB Layout for Router





A set of Router plugged on PC



Display for Video and Voice Conferencing





# Chapter 5

## Conclusion

The design and implementation of level-1 router for CUM LAUDE NET has been successfully demonstrated. The experimental result shows that it is possible to use non-expensive 486DX2-66 PC to build up a high speed hierarchical dual-ring multimedia network, including the high-speed network router, the hubs and the hosts.

Instead of supporting the ATM standard which is not yet widely available, CUM LAUDE NET has been designed and implemented to support high speed, real-time multimedia services with maximum compatibility to IP-based networks. In the future, when ATM service is available to the general public, the physical layer of CUM LAUDE NET can easily be changed to support ATM without any hardware modification.

We believe such an approach is very practical and feasible. As it stands now after only a year and a half's effort, CUM LAUDE NET has already been connected to Internet, and all standard Internet services like electronic mail (SMTP), remote login (TELNET), and file transfer (FTP) has been supported.



The user interface is the industry-standard X-windows. The network has been connected to the public telephone network through a T-1 gateway, thus allowing CUM LAUDE NET users to call up any telephone users and to send / receive voice mails or other customized services over a computer network.

A novel network protocol ACTA (Adaptive Cycle Tunable Access) has been implemented on CUM LAUDE NET. Fair access is achieved by limiting the number of empty slots occupied by each cycle. The cycle length is adjusted to reduce the packet latency and to increase the overall throughput. To increase the efficiency of the protocol implementation, the ACTA protocol has been slightly modified. Instead of generating a continuous stream of empty slots, a single access control packet containing the number of empty slots for different priority levels are used. This arrangement saves much packet processing time of the level-1 routers.

To summarize, the features of CUM LAUDE NET are as follows:

1. Hierarchical dual ring topology to provide single-fault tolerance.
2. 125 MBaud on each physical serial link with 100-Mbps data rate (level-1 routers and NIC).
3. Support variable-sized packets (fixed-size packet is used currently).
4. Packet format compatible with TCP/IP (ATM in the future).
5. Connect to public telephone network through a T-1 gateway.
6. Support real-time video and voice conferencing with 15-fps video.
7. Support voice mail, voice messaging, fax, and video on demand.

# Bibliography

- [1] R.M.Metcalf D.R.Boggs, "Ethernet: Distributed Packet Switching for Local Computer Networks," *Commun. ACM*, Vol. 19, pp.395-404, 1976.
- [2] W.S.Wai, "Packet mode services: from X.25 to frame relaying," *Computer Communications*, Vol. 12, No. 1, pp.10-16, Feb 1989.
- [3] "Special Issue on Digital Video and Multimedia," *Communications of the ACM*, April 1991.
- [4] "Special Issue on Interactive Multimedia," *IEEE Spectrum*, March 1993.
- [5] Borko Furht, "Multimedia Systems: An Overview," *IEEE Multimedia*, Spring, pp.47-59, 1994.
- [6] M.D.Prycker, *Asynchronous Transfer Mode: Solution for Broadband ISDN. Second Edition*. Ellis Horwood, 1993.
- [7] CCITT Recommendations, "CCITT Recommendations: G.702 G.703 G.704 G.706 G.707 G.708 G.709 I.113 I.121 I.150 I.211 I.311 I.321 I.327 I.361 I.362 I.363 I.364 I.371 I.413 I.414 I.430 I.431 I.432 I.441 I.600 I.610 M.20 M.30 M.36," *CCITT Recommendations*, 1992.



- [8] ATM Forum, "ATM User-Network Interface Specification Version 2.0," *ATM Forum*, Jun 1992.
- [9] F.E.Ross J.R.Hamstra, "Forging FDDI," *IEEE JSAC*, Vol. 11, No. 2, pp.181-190, Feb 1993.
- [10] W.Stallings, *Networking Standards: A Guide to OSI, ISDN, LAN & MAN Standard*. Addison-Wesley, 1993.
- [11] R.M.Needham A.J.Herbert, *The Cambridge Distributed Computing System*. Addison-Wesley, London, 1982.
- [12] G.Watson S.Ooi D.Skellern D.Cunningham, "Hangman Gb/s Network," *IEEE Network*, pp.10-18, July 1992.
- [13] K.W.Cheung C.T.Yeung W.K.Lam C.H.Chan C.W.Chung C.Lau K.K.Lau O.Soo C.K.Tong K.F.Wong K.F.Yuen, "CUM LAUDE NET - A High-Speed Multimedia Integrated Network Prototype," *ismm'94, Hawaii*, Aug 1994.
- [14] K.W.Cheung, "Adaptive-Cycle Tunable-Access (ACTA) Protocol: A Simple, High-Performance Protocol for Tunable-Channel Multi-Access (TCMA) Networks," *ICC'93, paper 16.1, pp.166-171, Geneva, Switzerland*, May 1993.
- [15] K.W.Cheung L.K.Chen C.Su C.T.Yeung P.T.To, "Tunable-Channel Multi-Access (TCMA) Networks: A New Class Of High-Speed Networks Suitable For Multimedia Integrated Networking," *SPIE'93 - Multigigabit Fiber Communication Systems, San Diego, CA*, July 1993.

- [16] J.O.Limb C.Flores, "Description of Fasnet — A Unidirectional Local-Area Communications Network," *Bell Syst. Tech. J.*, Vol. 61, pp.1413-1440, 1982.
- [17] ANSI, *FDDI Media Access Control (MAC-2)*. ANSI, X3T9.5/88-139, 1990.
- [18] F.E.Ross, "An Overview of FDDI: The Fiber Distributed Data Interface," *IEEE JSAC*, Vol. 7, pp.1043-1051, 1989.
- [19] R.M.Newman Z.L.Budrikis J.L.Hullett, "The QPSX MAN," *IEEE Communications Magazine*, pp.20-28, Apr. 1988.
- [20] Standard IEEE 802.6, "Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN)," *Standard IEEE 802.6*, Dec. 1990.
- [21] C.Partridge, *Gigabit Networking*. Addison-Wesley, 1993.
- [22] M.K.Johnson, *Linux Kernel Hackers' Guide*. GNU General Public License, 1992.
- [23] AMD, *Am7968/Am7969 TAXIchip Integrated Circuit User's Manual*. AMD, 1993.
- [24] AMD, *CMOS Memory Products*. AMD, 1991.
- [25] W.Rosch, "VL-Bus and PCI: Comparing the Local-Bus Standards," *PC Magazine*, pp.355 - 362, Oct 1993.
- [26] G.Blair A.Campbell G.Coulson, "A Network Interface Unit to Support Continuous Media," *IEEE JSAC*, Vol. 11, No. 2, Feb 1993.
- [27] Analog Devices, *ADSP-21000 Family User's Manual*. Analog Devices, 1993.





CUHK Libraries



000275738